

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:34:12 ; Search time 61.0859 seconds
(without alignments)
1246.205 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICOLKGAQMLCIDNCGA.....GRCLSQILHIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1599	100.0	475	11 Q8VIE0	Q8vie0 mus musculus
2	1589	99.4	475	11 Q8VIE1	Q8vie1 mus musculus
3	631	39.5	419	11 Q8C107	Q8c107 mus musculus
4	504	31.5	277	11 Q9D571	Q9d571 mus musculus
5	435	27.2	950	5 Q9V6Q8	Q9v6q8 drosophila
6	396	24.8	429	5 Q62295	Q62295 caenorhabdi
7	396	24.8	435	5 Q950D7	Q95qd7 caenorhabdi
8	363	22.7	571	10 Q9SL70	Q9sl70 arabidopsis
9	357	22.3	532	10 Q8G0Z0	Q8guz0 arabidopsis
10	357	22.3	536	10 Q8GUZ2	Q8guz2 arabidopsis
11	357	22.3	543	10 Q8GUZ1	Q8guz1 arabidopsis
12	357	22.3	603	10 Q9SD01	Q9sd01 arabidopsis
13	253.5	15.9	356	10 Q9LW71	Q9lw71 arabidopsis
14	251.5	15.7	601	10 Q9CAV1	Q9cav1 arabidopsis
15	232.5	14.5	896	10 Q9ZS22	Q9zs22 glycine max
16	232	14.5	658	10 Q23333	Q23333 arabidopsis

17	223.5	14.0	609	10 Q9M679	Q9m679 arabidopsis
18	223.5	14.0	609	10 Q8L548	Q8l548 arabidopsis
19	222	13.9	526	10 Q94A12	Q94a12 arabidopsis
20	222	13.9	526	10 Q8H1Q0	Q8h1q0 arabidopsis
21	222	13.9	695	10 Q9LUI3	Q9lui3 arabidopsis
22	222	13.9	695	10 Q9LE32	Q9lej2 arabidopsis
23	186	11.6	243	5 Q9VMQ3	Q9vmq3 drosophila
24	180.5	11.3	593	10 Q9LUI5	Q9lui5 arabidopsis
25	165	10.3	251	4 Q8NCC8	Q8ncc8 homo sapien
26	159	9.9	553	10 Q94DS2	Q94ds2 oryza sativ
27	118.5	7.4	1145	11 Q8BMN9	Q8bm9 mus musculu
28	118.5	7.4	1450	11 Q8JZW7	Q8jzw7 mus musculu
29	115	7.2	1737	4 Q75097	Q75097 homo sapien
30	105	6.6	4998	11 Q8CG65	Q8cg65 mus musculu
31	101	6.3	769	11 Q8VHL6	Q8vhl6 sigmodon hi
32	100	6.3	773	5 Q17514	Q17514 caenorhabdi
33	100	6.3	773	5 Q62335	Q62335 caenorhabdi
34	99.5	6.2	1704	5 Q94446	Q94446 chironomus
35	99	6.2	2841	5 Q8MLU9	Q8mlu9 drosophila
36	99	6.2	2931	5 Q9W2C6	Q9w2c6 drosophila
37	97.5	6.1	772	13 Q9PUU4	Q9puu4 ictalurus p
38	97	6.1	497	5 Q23460	Q23460 caenorhabdi
39	97	6.1	984	5 Q9Y1P7	Q9y1p7 cryptospori
40	97	6.1	1106	5 Q9GU55	Q9gu55 drosophila
41	97	6.1	2171	5 Q9V841	Q9v841 drosophila
42	96.5	6.0	1026	5 Q8SWY0	Q8swy0 drosophila
43	96.5	6.0	4547	5 Q9W343	Q9w343 drosophila
44	96	6.0	773	5 Q9VQR1	Q9vqr1 drosophila
45	95.5	6.0	448	5 Q9VJU8	Q9vju8 drosophila

ALIGNMENTS

RESULT 1

Q8VIE0 PRELIMINARY: PRT: 475 AA.
AC Q8VIE0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Tesmin (Hypothetical 50.6 kDa protein).
GN MTL5 OR TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB057423; BAB64935.1; -;
DR EMBL: BC024377; AAH24377.1; -;
DR MGD: MGI:1340029; Mtl5.
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 1.
KW Hypothetical protein.
SQ SEQUENCE 475 AA; 50615 MW; 4373EDI1188DCF79F CRC64;

Query Match 100.0%; Score 1599; DB 11; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.7e-164;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVICOLKGAQMLCIDNCGARELKALHLLPQYDQDSFFQSELPKMTTTLVGRLLPVPAK 60
|||||

Db 181 MVICOLKGAOMLCIDNCGARELKALHLLPOYDDQSSFPQSELKPMTTLVGRLLPVPK 240
 Qy 61 LNLITQVDNGALPSAVNGAAPPSPALOGPPKITYLGYCDGFCSSGDFCNSCNNLRHEL 120
 Db 241 LNLITQVDNGALPSAVNGAAPPSPALOGPPKITYLGYCDGFCSSGDFCNSCNNLRHEL 300
 Qy 121 ERFAKAIACLDNRPEAFOPKMGKGRGLGAALKRHSKGNCKRSGCLKNYCEYEAKEIMCSS 180
 Db 301 ERFAKAIACLDNRPEAFOPKMGKGRGLGAALKRHSKGNCKRSGCLKNYCEYEAKEIMCSS 360
 Qy 181 IKCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 240
 Db 361 IKCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 420
 Qy 241 EVWEATCACLLAOGEEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 Db 421 EVWEATCACLLAOGEEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 475

RESULT 2

Q8VIE1 ID Q8VIE1 PRELIMINARY; PRT; 475 AA.
 AC Q8VIE1;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DE 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Tesmin.
 GN TESM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
 RT "Structure of the tesmin gene encoding a testis-specific persistent
 RT protein, a possible multifunctional protein with dynamic changes of
 RT localization throughout spermatogenesis."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB057422; BAB64934.1;
 DR InterPro; IPR005172;
 DR Pfam; PF03638; CXG; 1.
 SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Query Match 99.4%; Score 1589; DB 11; Length 475;
 Best Local Similarity 99.3%; Pred. No. 2e-163;
 Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MVICOLKGAOMLCIDNCGARELKALHLLPOYDDQSSFPQSELKPMTTLVGRLLPVPK 60
 Db 181 MVICOLKGAOMLCIDNCGARELKALHLLPOYDDQSSFPQSELKPMTTLVGRLLPVPK 240
 Qy 61 LNLITQVDNGALPSAVNGAAPPSPALOGPPKITYLGYCDGFCSSGDFCNSCNNLRHEL 120
 Db 241 LNLITQVDNGALPSAVNGAAPPSPALOGPPKITYLGYCDGFCSSGDFCNSCNNLRHEL 300
 Qy 121 ERFAKAIACLDNRPEAFOPKMGKGRGLGAALKRHSKGNCKRSGCLKNYCEYEAKEIMCSS 180
 Db 301 ERFAKAIACLDNRPEAFOPKMGKGRGLGAALKRHSKGNCKRSGCLKNYCEYEAKEIMCSS 360
 Qy 181 IKCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 240
 Db 361 IKCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 420
 Qy 241 EVWEATCACLLAOGEEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 Db 421 EVWEATCACLLAOGEEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 475

RESULT 3

Q8C107 ID Q8C107 PRELIMINARY; PRT; 419 AA.

AC Q8C107;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Weakly similar to TESMIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RC MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK029288; BAC26373.1;
 SQ SEQUENCE 419 AA; 45301 MW; 24E3D4D454217D5B CRC64;
 Query Match 39.5%; Score 631; DB 11; Length 419;
 Best Local Similarity 44.3%; Pred. No. 1.2e-59;
 Matches 135; Conservative 44; Mismatches 77; Indels 49; Gaps 9;
 Qy 29 LPOYD-DQSSFPQSEL--PKPMTTLVGRLLPVPKLNLTQVDNGALPSAVNGAAPPSPG 85
 Db 113 LPOQPNLTNLPPTGTVLAPAGTGNVYAV-LPAQ--YVTLQOQSSVSIASNSFTGTS 169
 Qy 86 ALQPPKITTLSG-----YDCFSGDFCNSCSC-----NNLR 117
 Db 170 GIQTQARLPFNGIIPSESTSRPKPCNTKSLCLKYCDGFANGFCNCTNRYNLE 229
 Qy 118 HELERFAKAIACLDNRPEAFOPKMGKGRGLGAALKRHSKGNCKRSGCLKNYCEYEAKEIM 177
 Db 230 HENERQAKAIACLDNRPEAFOPKMGKGRGLGAALKRHSKGNCKRSGCLKNYCEYEAKEIM 289
 Qy 178 CSSICKCKIACKNYEESPERKMLMSTPHYMEPGDF-----ESSHYLSPAKFSPPK 227
 Db 290 CSSICKCKIACKNYEESPERKMLMSTPHYMEPGDF-----ESSHYLSPAKFSPPK 346
 Qy 228 LRK--NRQAFSCISWEVEATCACLLAOGEEAEQHCSPSLAEQMIIEFGRCLSQILHI 285
 Db 347 LNSAGGKLPFTFTVEAATCNCLLAQAEQADKKGKSKAAERMIIEFGRCLSMVINS 406
 Qy 286 EFKSK 290
 Db 407 AGRAK 411

RESULT 4

Q9D571 ID Q9D571 PRELIMINARY; PRT; 277 AA.
 AC Q9D571;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
 DE Adult male testis cDNA, RIKEN full-length enriched library,
 DE clone:4930509C02, full insert sequence.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,


```

RESULT 8
Q9SL70
ID Q9SL70 PRELIMINARY; PRT; 571 AA.
AC Q9SL70;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE At2g20110 protein.
GN AT2G20110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RC
RX MEDLINE=20083487; PubMed=10617197;

```

RA Fujii C., Mason T.M., Bowman C.L., Bainsdale W.E., Fendleyam I.V.,
RA Buell C.R., Ketchum K.A., Lee J.-J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrea A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RT Nature 402:761-768(1999).
RN [2]
RN Nature 402:761-768(1999).
RP
RP SEQUENCE FROM N.A.
RC
RC STRAIN=cv. Columbia;
RL
RL Lin x.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006081; AAD24386.1; -;
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
DR
DR SEQUENCE 571 AA; 61312 MW; 36701C5FA62B3B89 CRC64;
SQ

	Query Match	22.7%	Score 363;	DB 10; Length 571;
	Best Local Similarity	32.4%;	Pred. No. 1.9e-30;	
	Matches .81; Conservative	39; Mismatches	76; Indels	54; Gaps 8;
QY	36	SSPQQSELKPKMTLLVGRLL-----PVPAKLNLITQVD--NGALPSAVNQAFAF	81	
		: : : : :: : : : : : :		
Dd	46	SSMAAASIPSVIVTVTRPIITSQAPTVTATPPPPPQSOGIILHVPRIHRPESPNSMPR	105	
QY	82	PSPGALGGPPK-----ITLSCYGCDFFSGDFCNCSGC---NNLRHELEREKAIK	127	
		: : : : : : :		
Dd	106	PAGETRDGTGPQKKQCCKHRSRLKLYCEFCFASTGYCDGCNCVCNVFNNEVEPARQVAE	165	
QY	128	ACLDNRPEAFOPKMKGRLGAALK-----LRHSKCNCRKSGCLKNYCCEYEAKI	176	

Db	166	STLERPNAPFRPKIAASPHGGRRDNREVGDVVMLARHNKGCHCKKSGCLKKYCECFQANI	225
QY	177	MCSSICKCIACKNVEESPERRKMLMSTPH-----YME-----PGDFESSHYLSPAKFSFGP	225
Db	226	LCSENCKCLDCKNFEGSEYRQSLFHGEHSNHLAYLQHANAAITGATGSSGFAS-----APP	281
QY	226	PKLRKNRQAF	235
Db	282	PKRRKGQEIF	291
RESULT 9			
Q8GUZO			
ID	Q8GUZO	PRELIMINARY;	PRT; 532 AA.
AC	Q8GUZO;		
DT	01-MAR-2003	(TrEMBLrel. 23, Created)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Hypochemical protein (Fragment).		


```
GN T2G17.9.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao Y., Ishmael N., Kumar N., Redman J., Smith S., Riedmuller S.,
RA Utterback T., Whitelaw C.A., Fraser C.M., Town C.D.;
RT "Cloning and sequencing of full-length cDNAs for hypothetical genes
RT from chromosome 2 of Arabidopsis thaliana.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY163775; AAN85199.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 532 AA; 57033 MW; 78CF4E441175958D CRC64;

Query Match 22.3%; Score 357; DB 10; Length 532;
Best Local Similarity 32.0%; Pred. No. 7.9e-30;
Matches 80; Conservative 39; Mismatches 77; Indels 54; Gaps 8;

QY 36 SSFPOSELPPKMTTLVGRLL-----PVPAKLNLTQVD-NGALPSAVNGAAF 81
Db 11 SSMASASPSPITVTTRPIITSOAPTATPPIPPPPQSGIILHVPRIHRPSPNSMPR 70
QY 82 PSGPALQGGPK-----ITLSGYCDCFSSGDFCNSCSC-----NLRHHELERFKAIK 127
Db 71 PAGETRDGTPQKKKQCNCKHSRCLKLYCECFASGYTCDGCNCVCFNNVNEPARRQAVE 130
QY 128 ACLDRNPEAFQPKMGKGRGAAK-----LRHSGCNCCKRSGCLKNYCECYEAKI 176
Db 131 STLERPNPNAFRPKIAASPHGGRDNRREVGVDVVLARHNKGCHCKKSGCLKKYCECFQANI 190
QY 177 MCSSICKCIACKNYEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSOP 225
Db 191 LCSENCXCLDCKNFESEVRQSLFHGEHSHNLAYLQHANAATGAIGSSGFAS----APP 246
QY 226 PKLRKNRQAF 235
Db 247 PKRRKGQEIF 256

RESULT 10
QY 36 SSFPOSELPPKMTTLVGRLL-----PVPAKLNLTQVD-NGALPSAVNGAAF 81
Db 11 SSMASASPSPITVTTRPIITSOAPTATPPIPPPPQSGIILHVPRIHRPSPNSMPR 70
QY 82 PSGPALQGGPK-----ITLSGYCDCFSSGDFCNSCSC-----NLRHHELERFKAIK 127
Db 71 PAGETRDGTPQKKKQCNCKHSRCLKLYCECFASGYTCDGCNCVCFNNVNEPARRQAVE 130
QY 128 ACLDRNPEAFQPKMGKGRGAAK-----LRHSGCNCCKRSGCLKNYCECYEAKI 176
Db 131 STLERPNPNAFRPKIAASPHGGRDNRREVGVDVVLARHNKGCHCKKSGCLKKYCECFQANI 190
QY 177 MCSSICKCIACKNYEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSOP 225
Db 191 LCSENCXCLDCKNFESEVRQSLFHGEHSHNLAYLQHANAATGAIGSSGFAS----APP 246
QY 226 PKLRKNRQAF 235
Db 247 PKRRKGQEIF 256

RESULT 11
QY 36 SSFPOSELPPKMTTLVGRLL-----PVPAKLNLTQVD-NGALPSAVNGAAF 81
Db 11 SSMASASPSPITVTTRPIITSOAPTATPPIPPPPQSGIILHVPRIHRPSPNSMPR 70
QY 82 PSGPALQGGPK-----ITLSGYCDCFSSGDFCNSCSC-----NLRHHELERFKAIK 127
Db 71 PAGETRDGTPQKKKQCNCKHSRCLKLYCECFASGYTCDGCNCVCFNNVNEPARRQAVE 130
QY 128 ACLDRNPEAFQPKMGKGRGAAK-----LRHSGCNCCKRSGCLKNYCECYEAKI 176
Db 131 STLERPNPNAFRPKIAASPHGGRDNRREVGVDVVLARHNKGCHCKKSGCLKKYCECFQANI 190
QY 177 MCSSICKCIACKNYEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSOP 225
Db 191 LCSENCXCLDCKNFESEVRQSLFHGEHSHNLAYLQHANAATGAIGSSGFAS----APP 246
QY 226 PKLRKNRQAF 235
Db 247 PKRRKGQEIF 256

RESULT 12
QY 36 SSFPOSELPPKMTTLVGRLL-----PVPAKLNLTQVD-NGALPSAVNGAAF 81
Db 11 SSMASASPSPITVTTRPIITSOAPTATPPIPPPPQSGIILHVPRIHRPSPNSMPR 70
QY 82 PSGPALQGGPK-----ITLSGYCDCFSSGDFCNSCSC-----NLRHHELERFKAIK 127
Db 71 PAGETRDGTPQKKKQCNCKHSRCLKLYCECFASGYTCDGCNCVCFNNVNEPARRQAVE 130
QY 128 ACLDRNPEAFQPKMGKGRGAAK-----LRHSGCNCCKRSGCLKNYCECYEAKI 176
Db 131 STLERPNPNAFRPKIAASPHGGRDNRREVGVDVVLARHNKGCHCKKSGCLKKYCECFQANI 190
QY 177 MCSSICKCIACKNYEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSOP 225
Db 191 LCSENCXCLDCKNFESEVRQSLFHGEHSHNLAYLQHANAATGAIGSSGFAS----APP 246
QY 226 PKLRKNRQAF 235
Db 247 PKRRKGQEIF 256
```

ID	Q9LW71	PRELIMINARY:	PRT;	356 AA.				
AC	Q9LW71;							
DT	01-OCT-2000	(TrEMBLrel. 15, Created)						
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)						
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)						
DE	Genomic DNA, chromosome 3, pl clone: MSLI.							
OS	Arabidopsis thaliana (Mouse-ear cress).							
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;							
OX	eurosid II; Brassicales; Brassicaceae; Arabidopsi.							
NCBI_TaxID=3702;								
[1]								
RN	SEQUENCE FROM N.A.							
RP	STRAIN=Columbia;							
RC	Sato S.; Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;							
RL	Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.							
[2]								
RN	SEQUENCE FROM N.A.							
RP	STRAIN=Columbia;							
RC	MEDLINE=20277480; PubMed=10819329;							
RX	Nakamura Y.;							
RA	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence							
RT	features of the regions of 4,504,864 bp covered by sixty pl and TAC							
RT	clones.";							
RL	DNA Res. 7:131-135(2000).							
DR	EMBL; AB012247; BAB02682.1; -.							
DR	InterPro; IPR005172; CXC.							
DR	InterPro; IPR000345; CytC_heme_bind.							
DR	pfam; PF03638; CXC; 2.							
DR	PROSITE; PS00190; CYTOCHROME_C; 1.							
SQ	SEQUENCE 356 AA; 39796 MW; 4FE0EA2D9DBFF58 CRC64;							
<hr/>								
Query Match 15.9%; Score 253.5; DB 10; Length 356;								
Best Local Similarity 32.1%; Pred. No. 7.8e-19;								
Matches 77; Conservative 30; Mismatches 90; Indels 43; Gaps 11								
QY	98	YDCFPSSGDFCNSSCNNLRHELRFKAIKAC----	LDRNPFAFQPMKGKGRIGAAKLRH	153				
Db	79	YDCDFASGVCTDCDCVCDHNNSEKCDAREAAAMVNVLGRPNPAFSEK-	ALGSLTDNQCA	137				
QY	154	-----SKGCNCKRSGCLNKVCCEYAKIMCSSICKCIACKNYEES--PERKMLMST	202					
Db	138	APDKPGLLSRGCKRTRCLKKYECFQANLLCSNDCKINCINKNSEAFQPPAFAHNS	197					
QY	203	PH-YMEPGDFESSHYLS-PAK-FSGP-----PKL--RKNRQAFSCISWEV	242					
Db	198	PQVYRRRRDRLETWNSCPAPLFSPIDNSIONALGSPMSCSPKLYPKKKRLMGYSTLL	257					
QY	243	VEA---TCACLAAQGEA-----EQEHCSPSLAQMILEEPFGCLSQLIHIEPKSGLKTE	295					
Db	258	PDLGDLCSLLVAASEATTTAEDQNRIFTFKPDOKAEITELSESSESNRVEEIQSRGLIE	317					

198	Db		POVRRRRDRRETTBNWSCPAPLFI	PSPDNSTONALGSPMCSCKPLPYRKRKSLRMGYTSTLL	257
243	QY	VEA--TCACLLAQGEA-----E	QEHCSPSLAEOMILEEFGRCLSQILHIEFKSKGLKTE		295
258	Db		PDLGDCLSLLVAASESATTTAEDONRI	FTRPDDKEAIELSESESERNVEEETQSGRLLIE	317
RESULT 14					
Q9CAV1					
ID		Q9CAV1	PRELIMINARY;	PRT;	601 AA.
AC		Q9CAV1:			
DT		01-JUN-2001	(TREMBLrel. 17, Created)		
DT		01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT		01-OCT-2002	(TREMBLrel. 22, Last annotation update)		
DE		Hypothetical 66.8 kda protein.			
DE		T9J14.20.			
GN		Arabidopsis thaliana (Mouse-ear cress).			
OS		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC		eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
NCBI_TaxID=3702;		[1]			
OX		SEQUENCE FROM N.A.			
RX		STRAIN=cv. Columbia;			
RC		MEDLINE=21016720; PubMed=11130713;			
RX		Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,			
RA		Partmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,			

RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA MEDLINE=21016740; PUBMED=11130713;

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:22:37 ; Search time 14.899 Seconds
(without alignments)
931.129 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLIHIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1599	100.0	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
2	1251	78.2	299	1 MTL5_HUMAN	Q9y415 homo sapien
3	110.5	6.9	971	1 RECK_HUMAN	Q95980 homo sapien
4	103.5	6.5	643	1 CD93_RAT	Q9et61 rattus norv
5	100	6.3	773	1 MES2_CAEEL	P17514 caenorhabdi
6	96.5	6.0	1790	1 LMB1_DROME	P11014 drosophila
7	96	6.0	773	1 MS12_DROME	P50334 drosophila
8	96	6.0	3718	1 LMA5_MOUSE	Q61001 mus musculus
9	95	5.9	1096	1 KPCL_ASPNG	Q00078 aspergillus
10	93.5	5.8	652	1 CD93_HUMAN	Q9npy3 homo sapien
11	93	5.8	1816	1 LMA4_HUMAN	Q16363 homo sapien
12	92	5.8	798	1 ITB1_XENLA	P12606 xenopus lae
13	91.5	5.7	644	1 CD93_MOUSE	Q89103 mus musculus
14	91	5.7	493	1 OC90_HUMAN	Q02509 homo sapien
15	91	5.7	747	1 EZH1_HUMAN	Q92800 homo sapien
16	91	5.7	2907	1 FBN2_MOUSE	Q61555 mus musculus
17	90.5	5.7	735	1 AD02_MACFA	Q28478 macaca fasc
18	89.5	5.6	459	1 RN25_HUMAN	Q96bh1 homo sapien
19	89	5.6	1786	1 LMB1_MOUSE	P02469 mus musculus
20	88	5.5	798	1 ITB0_XENLA	P12607 xenopus lae
21	88	5.5	1786	1 LMB1_HUMAN	P07942 homo sapien
22	87.5	5.5	971	1 RECK_MOUSE	Q9z0j1 mus musculus
23	87	5.4	290	1 Y070_NPVAC	P41470 autographa
24	87	5.4	501	1 TRA2_HUMAN	Q12933 homo sapien
25	87	5.4	652	1 SYG_SCHPO	Q10179 schizosacch
26	87	5.4	902	1 CLF_ARATH	P93831 arabidopsis
27	86.5	5.4	611	1 LEM2_CANFA	P33730 canis fami
28	86	5.4	518	1 KCG2_CHICK	Q73606 gallus gall
29	86	5.4	610	1 LEM2_HUMAN	P16581 homo sapien
30	86	5.4	747	1 EZH1_MOUSE	P70351 mus musculus
31	86	5.4	830	1 LEM3_HUMAN	P16109 homo sapien
32	85.5	5.3	769	1 ITB2_HUMAN	P05107 homo sapien
33	85.5	5.3	806	1 ITB7_MOUSE	P26011 mus musculus

34	85	5.3	398	1 MUB1_XENLA	P38565 xenopus lae
35	85	5.3	746	1 EZH2_MOUSE	Q61188 mus musculus
36	85	5.3	787	1 ITB6_MOUSE	Q9z0e9 mus musculus
37	84.5	5.3	169	1 KRUA_HUMAN	P26371 homo sapien
38	84.5	5.3	720	1 NRG3_HUMAN	P56975 homo sapien
39	84.5	5.3	1639	1 LMGI_DROME	P15215 drosophila
40	84.5	5.3	1816	1 LMA4_MOUSE	P97927 mus musculus
41	84	5.3	769	1 ITB2_BOVIN	P32592 bos taurus
42	84	5.3	2531	1 NTC1_MOUSE	Q01705 mus musculus
43	84	5.3	2531	1 NTC1_MOUSE	Q07008 rattus norv
44	84	5.3	3707	1 PGBM_RAT	Q05793 mus musculus
45	84	5.3	3828	1 TRX_DROVI	Q24742 drosophila

ALIGNMENTS

RESULT 1
MTL5_MOUSE STANDARD; PRT; 295 AA.
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).
GN MTL5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis; PubMed=10191092;
RX MEDLINE=99208689; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77383; AAD24667.1; -;
DR EMBL; U67176; AAD24666.1; -;
DR MGD; MGI:1340029; Mtl5.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Spermatogenesis.
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Query Match					100.0%; Score 1599; DB 1; Length 295;
Best Local Similarity					100.0%; Pred. No. 2.1e-130;
Matches 295; Conservative					0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVICQLKGAQMLCIDNCGARELKALHLLPQYDQSSFFQSELPKPMTLVGRLLPVPK 60					
-Db 1 MVICQLKGAQMLCIDNCGARELKALHLLPQYDQSSFFQSELPKPMTLVGRLLPVPK 60					
QY 61 LNLITQVDNGALPSAVNGAAPPSPGALQGPPTLTSGYDCDFSSGDFCNSCNSNLRHEL 120					
Db 61 LNLITQVDNGALPSAVNGAAPPSPGALQGPPTLTSGYDCDFSSGDFCNSCNSNLRHEL 120					

```
QY 121 ERFAIKACLDNRPEAFQPKMGKGLGAALKRHSKGNCRSGGLKNKYCEYEAQIMCSS 180
|||||
Db 121 ERFAIKACLDNRPEAFQPKMGKGLGAALKRHSKGNCRSGGLKNKYCEYEAQIMCSS 180
|||||
QY 181 IKCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPPKLRKQAFSCISW 240
|||||
Db 181 IKCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPPKLRKQAFSCISW 240
|||||
QY 241 EWEATACLAQGEAEQEHCSPLABQMILEEFGRCLSQILHIEFKSKGLKIE 295
|||||
Db 241 EWEATACLAQGEAEQEHCSPLABQMILEEFGRCLSQILHIEFKSKGLKIE 295
|||||

RESULT 2
MTL5_HUMAN
ID MTL5_HUMAN STANDARD; PRT; 299 AA.
AC Q9V4I5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RT early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U86074; AAD24668.1; -.
DR Genew; HGNC:7446; MTL5.
DR MIM; 604374; -.
DR GO; GO:0005505; F:heavy metal binding activity; TAS.
DR GO; GO:0006875; P:metal ion homeostasis; TAS.
DR GO; GO:0009412; P:response to heavy metal; TAS.
DR GO; GO:0007283; P:spermatogenesis; TAS.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH.
SQ SEQUENCE 299 AA; 33110 MW; 8F679F140F969C11 CRC64;

Query Match 78.2%; Score 1251; DB 1; Length 299;
Best Local Similarity 76.3%; Pred. No. 1.8e-100;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

QY 1 MVICQKGAOMCIDNCGARELKAHLHPYDQSSFPQSELKPKMTTLVGRLLPVPK 60
|||||
Db 1 MVICQKGTQMLCIDNSTRRELKAHLVLPQYQDNLYLQSDVFKPMFALVGRFLPASTK 60
|||||
QY 61 LNLITQVNDGALPSAVNGAARFSPGALQGPPIKILSGYCDGFCSSGDCNCS----CNL 116
|||||
Db 61 LNLITQLEGALPSVYNGAARFSPGSTLPKPKITLAGYCDGFCFASGDFCNCCNCCNL 120
|||||
QY 117 RHELFERKAIACLDNRPEAFQPKMGKGLGAALKRHSKGNCRSGGLKNKYCEYEAQI 176
|||||
```

```
Db 121 HHDIERFAIKACLDNRPEAFQPKMGKGLGNVRPQHNGKNCRRSGGLKNKYCEYEAQI 180
|||||
QY 177 MCSSICICACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPPKLRKQAFS 236
|||||
Db 181 MCSSICICACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPPKLRKQAFS 240
|||||
QY 237 CISWEVVEATCACLQAQGEAEQEHCSPLAEQMILEEFGRCLSQILHIEFKSKGLKIE 295
|||||
Db 241 CISWEVVEATCACLQAQGEAEQEHCSPLAEQMILEEFGRCLSQILHIEFKSKGLKIE 299
|||||

RESULT 3
RECK_HUMAN
ID RECK_HUMAN STANDARD; PRT; 971 AA.
AC O95980; O8WX37;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Reversion-inducing cysteine-rich protein with Kazal motifs precursor
DE (hRECK) (Suppressor of tumorigenicity 15) (ST15).
GN RECK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND N-GLYCOSYLATION.
RC TISSUE=Fibroblast;
RX MEDLINE=99007295; PubMed=9789069;
RA Takahashi C., Sheng Z., Horan T.P., Kitayama H., Maki M., Hitomi K.,
RA Kitaura Y., Takai S., Sasahara R.M., Horimoto A., Ikawa Y.,
RA Ratzkin B.J., Arakawa T., Noda M.;
RT "Regulation of matrix metalloproteinase-9 and inhibition of tumor
RT invasion by the membrane-anchored glycoprotein RECK.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:13221-13226(1998).
RN [2]
RP SEQUENCE OF 363-971 FROM N.A.
RA Kimberley A.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Negatively regulates matrix metalloproteinase-9 (MMP-9)
CC by suppressing MMP-9 secretion and by direct inhibition of its
CC enzymatic activity. RECK down-regulation by oncogenic signals may
CC facilitate tumor invasion and metastasis. Appears to also
CC regulate MMP-2 and MT1-MMP, which are involved in cancer
CC progression.
CC -!- SUBUNIT: Interacts with MMP-9.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- TISSUE SPECIFICITY: Expressed in various tissues and untransformed
CC cells. It is undetectable in tumor-derived cell lines and
CC oncogenically transformed cells.
CC -!- SIMILARITY: Contains 3 Kazal-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50406; BAA34060.1; -.
DR EMBL; AL158830; CADI3384.1; -.
DR Genew; HGNC:11345; RECK.
DR MIM; 605227; -.
DR HSP; P80424; 1AN1.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0008191; F:metalloendopeptidase inhibitor activity; TAS.
DR InterPro; IPR002350; kazal.
DR Pfam; PF00050; kazal; 2.
DR SMART; SM00280; KAZAL; 3.
DR PROSITE; PS00282; KAZAL; 1.
KW Signal; Glycoprotein; GPI-anchor; Serine protease inhibitor;
```


FT DISULFID 308 325 BY SIMILARITY.
FT DISULFID 327 340 BY SIMILARITY.
FT DISULFID 346 355 BY SIMILARITY.
FT DISULFID 351 364 BY SIMILARITY.
FT DISULFID 366 380 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 393 406 BY SIMILARITY.
FT DISULFID 408 422 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 433 446 BY SIMILARITY.
FT DISULFID 448 461 BY SIMILARITY.
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 417 417 E -> K (IN REF. 2).
SQ SEQUENCE 643 AA; 68781 MW; 9AE4C933AD943DB6 CRC64;

Query Match 6.5%; Score 103.5; DB 1; Length 643;
Best Local Similarity 29.4%; Pred. No. 0.21;
Matches 37; Conservative 14; Mismatches 54; Indels 21; Gaps 6;

Qy 83 SGPALQPPKTTLS---GYC--DCFSSGDFCNCSNNLRHELFRKAKACLDNRPEAF 137
Db 250 SGP-LCVSPKFCGCFNNGGCOOCFEGGDSFRGCG---RPGFRLDDLVTCASRNPCCS 305

Qy 138 QPKMGKGRGAAKLRHSGKNCXRGGLKNVCYEYKIMCSCICKIACKNYEESPERK 197
Db 306 NPCTGGGGMCHSVLSENYTCHPCRGYQLDS-----SQVHCVDIDEC-----EDSPCDQ 353

Qy 198 MLMSTP 203
Db 354 ECINTP 359

RESULT 5

MES2_CABEL STANDARD; PRT: 773 AA.
AC 017514; O62335;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polycumb protein mes-2 (Maternal-effect sterile 2 protein) (E(z)
DE homolog)
DE MES-2 OR R06A4.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, DEVELOPMENTAL STAGE, AND MUTANTS
RC BN48 AND BN72.
RC STRAIN-Bristol N2;
RX MEDLINE=98274205; PubMed=9609829;
RA Holdeman R., Nehrt S., Strome S.;
RT "MES-2, a maternal protein essential for viability of the germline in
RT Caenorhabditis elegans, is homologous to a Drosophila Polycumb group
RT protein.";
RL Development 125:2457-2467(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Durbin R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP IDENTIFICATION IN A COMPLEX WITH MES-3 AND MES-6.
RX MEDLINE=21221078; PubMed=11320248;
RA Xu L., Fong Y., Strome S.;
RT "The Caenorhabditis elegans maternal-effect sterile proteins, MES-2,
RT MES-3, and MES-6, are associated in a complex in embryos.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:5061-5066(2001).

Query Match 6.3%; Score 100; DB 1; Length 773;
Best Local Similarity 21.8%; Pred. No. 0.52;
Matches 32; Conservative 14; Mismatches 41; Indels 60; Gaps 7;

Qy 89 GPKKITLSGYCDGFCSSGDFCNCSNNLRHELFRKAKACLDNRPEAFQPKMGKGRUGA 148
Db 535 GPCNATAEN-CACRENGVCSYMKCD-----INCSQRF----- 567
Qy 149 AKLRHSGKNCXRGGLKNVCYEYKIMCS-----SICKC-----IACKNYEESPERK 197

FN FUNCTION.
RX MEDLINE=22072749; PubMed=12077420;
RA Fong Y., Bender L., Wang W., Strome S.;
RT "Regulation of the different chromatin states of autosomes and X
RL chromosomes in the germ line of C. elegans.";
Science 296:2235-2238(2002).
CC -!- FUNCTION: Polycumb group (PCG) protein. PCG proteins act by
CC forming multiprotein complexes, which are required to maintain the
CC transcriptionally repressive state of homeotic genes throughout
CC development. PCG proteins are not required to initiate repression,
CC but to maintain it during later stages of development. The mes-
CC 2/mes-3/mes-6 complex may participate in the global inactivation
CC of the X chromosomes in germ line cells. The complex may act via a
CC methylation of histones, rendering chromatin heritably changed in
CC its expressibility. Probable catalytic subunit of a potential
CC histone methyltransferase activity of the complex. This complex is
CC required to exclude mes-4 from the inactivated X-chromosomes in germ
CC line cells.
CC -!- SUBUNIT: Interacts directly with mes-6 via its N-terminal domain.
CC Forms a heterotrimeric complex with mes-3 and mes-6. Does not
CC interact with mes-4.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: In adults, it is predominantly expressed in
CC the germline, and weakly expressed in intestinal cells.
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC Expressed in all cells of early embryos. In late embryos and L1
CC larva, it is weakly expressed, while it is expressed at
CC intermediate levels in the germline of L4 larvae.
CC -!- SIMILARITY: BELONGS TO THE EZ FAMILY.
CC -!- SIMILARITY: Contains 1 SET domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF011893; AAC27124.1; -
CC EMBL; Z83120; CAB05589.2; -
CC EMBL; Z81515; CAB05589.2; JOINED.
CC EMBL; Z81515; CAB04199.2; -
CC EMBL; Z83120; CAB04199.2; JOINED.
CC WormPep; R06A4.7; CE28067.
CC InterPro; IPR01214; SET.
CC Pfam; PF00856; SET; 1.
CC SMART; SM00317; SET; 1.
CC PROSITE; PS0280; SET; 1.
CC Transcription regulation; Repressor; Nuclear protein;
CC Developmental protein.
CC DOMAIN 1 194 INTERACTION WITH MES-6.
CC DOMAIN 531 600 CYS-RICH.
CC DOMAIN 620 741 SET.
CC MUTAGEN 628 629 SK->PE: IN BN48; MATERNAL-EFFECT
CC MUTAGEN 674 674 MUTATION. PROGENY DEFECTS IN GONAD
CC PROLIFERATION. GERM CELL DEGENERATION.
CC Y->H: IN BN72; MATERNAL-EFFECT MUTATION.
CC PROGENY DEFECTS IN GONAD PROLIFERATION.
CC GERM CELL DEGENERATION.
CC CONFLICT 483 484 RK -> VO (IN REF. 1).
SQ SEQUENCE 773 AA; 88820 MW; 91ABEDAD94A1D51E CRC64;


```

Db 568 -----GCNCAAGOCYTRKACOCYCRANWECNPMTCMCKDAIDSNIIKRNFGMT--R 617
QY 198 MLMSTPHYMEPGDFESSHYLSPAKESG 224
Db 618 MI-----QKRTYCGPSKIAG 632

RESULT 6
LMB1_DROME
ID LMB1_DROME STANDARD: PRT: 1790 AA.
AC P11046; Q26328; Q9VLW6; Q9X2T4;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LAMB1 OR LAMB1 OR CG7123.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephedroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=94000382; PubMed=8397815;
RA Montell D.J., Goodman C.S.;
RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain
RT reveals domains of homology with mouse.";
RL Cell 53:463-473(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Deicher S., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster G., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

```

```

Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC -!- SIMILARITY: Contains 13 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 2 laminin IV domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M95811; AAD19752.1; -
CC EMBL: M19525; AAZ28663.1; -
CC EMBL: AE003618; AAF52563.1; -
CC PIR: A28783; MWFEB1.
CC HSP: P02468; LTLE.
CC FlyBase: FBgn0002527; LanB1.
CC GO: GO:0005605; C:basal lamina; IDA.
CC InterPro: IPR006209; EGF_like.
CC InterPro: IPR002049; Laminin_EGF.
CC Pfam: PF00053; laminin_EGF; 13.
CC Pfam: PF00055; laminin_Nterm; 1.
CC PRINTS: PR00011; EGF_LAMININ.
CC SMART: SM00180; EGF_Lam; 12.
CC SMART: SM00136; LamNT; 1.
CC PROSITE: PS00022; EGF_1; 10.
CC PROSITE: PS01186; EGF_2; 2.
CC PROSITE: PS01248; LAMININ_TYPE_EGF; 12.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL 1 25
CC CHAIN 26 1790 LAMININ BETA-1 CHAIN.
CC DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).
CC DOMAIN 290 356 LAMININ EGF-LIKE 1.
CC DOMAIN 357 419 LAMININ EGF-LIKE 2.
CC DOMAIN 420 479 LAMININ EGF-LIKE 3.
CC DOMAIN 480 530 LAMININ EGF-LIKE 4.
CC DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
CC DOMAIN 562 789 LAMININ EGF-LIKE 6.
CC DOMAIN 791 838 LAMININ EGF-LIKE 7.
CC DOMAIN 839 884 LAMININ EGF-LIKE 8.
CC DOMAIN 885 934 LAMININ EGF-LIKE 9.
CC DOMAIN 935 992 LAMININ EGF-LIKE 10.
CC DOMAIN 993 1044 LAMININ EGF-LIKE 11.
CC DOMAIN 1045 1095 LAMININ EGF-LIKE 12.
CC DOMAIN 1096 1143 LAMININ EGF-LIKE 13.
CC DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
CC DOMAIN 1191 1407 DOMAIN II.
CC DOMAIN 1408 1434 DOMAIN ALPHA.
CC DOMAIN 1435 1790 DOMAIN I.
CC DOMAIN 1257 1407 COILED COIL (POTENTIAL).

```



```

FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
FT DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
FT DOMAIN 2169 2735 DOMAIN II AND I.
FT DOMAIN 2736 2933 LAMININ G-LIKE 1.
FT DOMAIN 2934 3119 LAMININ G-LIKE 2.
FT DOMAIN 3120 3296 LAMININ G-LIKE 3.
FT DOMAIN 3297 3511 LAMININ G-LIKE 4.
FT DOMAIN 3512 3689 LAMININ G-LIKE 5.
FT DOMAIN 3690 4100 COILED COIL (POTENTIAL).
FT DOMAIN 4101 431 LAMININ G-LIKE 5.
FT DOMAIN 432 445 COILED COIL (POTENTIAL).
FT DOMAIN 446 452 COILED COIL (POTENTIAL).
FT DOMAIN 453 463 COILED COIL (POTENTIAL).
FT DOMAIN 464 476 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 477 512 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 513 521 BY SIMILARITY.
FT DOMAIN 522 532 BY SIMILARITY.
FT DOMAIN 533 544 BY SIMILARITY.
FT DOMAIN 545 559 BY SIMILARITY.
FT DOMAIN 560 566 BY SIMILARITY.
FT DOMAIN 567 577 BY SIMILARITY.
FT DOMAIN 578 590 BY SIMILARITY.
FT DOMAIN 591 605 BY SIMILARITY.
FT DOMAIN 606 611 BY SIMILARITY.
FT DOMAIN 612 622 BY SIMILARITY.
FT DOMAIN 623 635 BY SIMILARITY.
FT DOMAIN 636 650 BY SIMILARITY.
FT DOMAIN 651 656 BY SIMILARITY.
FT DOMAIN 657 667 BY SIMILARITY.
FT DOMAIN 668 680 BY SIMILARITY.
FT DOMAIN 681 695 BY SIMILARITY.
FT DOMAIN 696 702 BY SIMILARITY.
FT DOMAIN 703 713 BY SIMILARITY.
FT DOMAIN 714 726 BY SIMILARITY.
FT DOMAIN 727 1455 BY SIMILARITY.
FT DOMAIN 1456 1462 BY SIMILARITY.
FT DOMAIN 1463 1473 BY SIMILARITY.
FT DOMAIN 1474 1486 BY SIMILARITY.
FT DOMAIN 1487 1533 BY SIMILARITY.
FT DOMAIN 1534 1555 BY SIMILARITY.

FT DISULFID 1557 1566 BY SIMILARITY.
FT DISULFID 1567 1579 BY SIMILARITY.
FT DISULFID 1580 1594 BY SIMILARITY.
FT DISULFID 1595 1601 BY SIMILARITY.
FT DISULFID 1602 1612 BY SIMILARITY.
FT DISULFID 1613 1630 BY SIMILARITY.
FT DISULFID 1631 1674 BY SIMILARITY.
FT DISULFID 1675 1881 BY SIMILARITY.
FT DISULFID 1882 1893 BY SIMILARITY.
FT DISULFID 1894 1912 BY SIMILARITY.
FT DISULFID 1913 1930 BY SIMILARITY.
FT DISULFID 1931 1939 BY SIMILARITY.
FT DISULFID 1940 1950 BY SIMILARITY.
FT DISULFID 1951 1968 BY SIMILARITY.
FT DISULFID 1969 1986 BY SIMILARITY.
FT DISULFID 1987 1993 BY SIMILARITY.
FT DISULFID 1994 2005 BY SIMILARITY.
FT DISULFID 2006 2022 BY SIMILARITY.
FT DISULFID 2023 2083 BY SIMILARITY.
FT DISULFID 2084 2090 BY SIMILARITY.
FT DISULFID 2091 2101 BY SIMILARITY.
FT DISULFID 2102 2116 BY SIMILARITY.
FT DISULFID 2117 2126 BY SIMILARITY.
FT DISULFID 2127 2133 BY SIMILARITY.
FT DISULFID 2134 2144 BY SIMILARITY.
FT DISULFID 2145 2166 BY SIMILARITY.
FT DISULFID 2167 2169 INTERCHAIN (PROBABLE).
FT DISULFID 2170 2172 INTERCHAIN (PROBABLE).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 6.0%; Score 96; DB 1; Length 3718;
Best Local Similarity 18.0%; Pred. No. 6.9;
Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;

Qy 14 CIDNCGARELKALHLLPOYDDQSFQSELPKPMWTLVGLLPPAKNLITQVDNCALP 73
Db 200 CLERFGPRTLER---ITQDDVI-----CTTEYSRIVPL-ENGEIVSVLVNGR-P 244
Qy 74 SAVNGAAPP-----SGPALQGP-----PKITLSGY 98
Db 245 GALNFSYSPILLRDTFKATNIRLRLTNTLLGLHLMGKALRDPVTRRYYSIKDISIGR 304
Qy 99 CDFSSGDFC-----NSC-----SC 113
Db 305 CVCHGHADVCDKADPLDPFRLOACQHCQDCRCGPGFNQOPWKPATTDSANECQSC 364
Qy 114 NNLEH-----ELERFKAICACLDNRNPAFQPKMGKRLGAALKL---RHSKGCNCKRS 162
Db 365 NCHGHAYDCYYDPEVDR-----RNASONQDNVYQG--GGVCLDCQHHTTGINGER- 412
Qy 163 GCLKNY-----CECYEAKI---MCSSTI---CK-----CIAC-KNY 190
Db 413 -CLPGFRAPDQPLDPSHVCRPCDC-ESDRTDGTCDLTGRCYCRPNFTGELCAACAGY 470
Qy 191 EESPERKMLMSTHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISWEWEATCACL 250
Db 471 TDFPHCYPLPSFPHN-----DTRQVLP-----QIVNCDNCAA 505
Qy 251 LAQGEAEQE-----HC-----SPSLAEQMLLEFGR 277
Db 506 GTQGNACRKDPRLGRVCVKPNFRGAHCELCAPEGHSPCHPCQCSGPGVANSLCDPESGQ 565
Qy 278 CL 279
Db 566 CM 567

RESULT 9
KPC1 ASPNG
ID KPC1 ASPNG STANDARD; PRT; 1096 AA.
AC Q00078;
DT 01-NOV-1997 (Rel. 35, Created)
```

QY	121	ERFKAI-----KACIDRNPDEAFQPMKGKGRGLGAALKUR-----HSKG-----	157
DB	416	EPVGRIVYLSMSFAKQLKDRRF-----DIGLNROQAVRKGKEVHKQGHKFTVQGFYNIM	471
QY	158	NKRSRSGCLKNVC-----ECVEAKIMCSSIC-----KCIACKNYEESPERKMLMSTHYWEP	208
DB	472	RCALCGDFLKYAGMQCADCKYTKHKCYKPVVTKICISANYETDP-----	517
QY	209	GDPESSHYLSPAKSGPPKLRKNRQAQSCISWEVVEATCACCILAQGEAEQHCSPSLAE	268
DB	518	DEEKINHRIPHREEG-----FSNIS-----ANWCC-----HCG-----	545
QY	269	OMILEEPR-----CLSQILHIEFKSKGLKTE	295
DB	546	--YLLPGRKNAKRCETCGLTCHSCHTHLVDFCGMSME	582

RESULT 10

CD93_HUMAN

ID

CD93_HUMAN

STANDARD;

PRT;

652 AA.

AC

Q9NPY3; O00274;

DT

28-FEB-2003 (Rel. 41, Created)

DT

28-FEB-2003 (Rel. 41, Last sequence update)

DT

15-SEP-2003 (Rel. 42, Last annotation update)

DE

Complement component C1q receptor precursor (Complement component 1, q subcomponent, receptor 1) (C1qRp) (C1qR(p)) (C1q/MBL/SPA receptor) (CD93 antigen) (CDW93).

DE

C1QR1 OR CD93.

GN

Human.

OS

Homo sapiens (Human).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX

NCBI_TaxId=9606;

RI

[1]

RN

SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX

MEDLINE=97199258; PubMed=9047234;

RA

Nepomuceno R.R., Henschen-Edman A.H., Burgess W.H., Tenner A.J.; "cDNA cloning and primary structure analysis of C1qR(P), the human C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro."; Immunity 6:119-129(1997).

RI

[2]

RN

SEQUENCE FROM N.A., AND VARIANT ALA-318.

RP

MEDLINE=21640567; PubMed=11781389;

RX

Steinberger P., Szekeres A., Wille S., Stockl J., Selenko N., Prager E., Staffler G., Medic O., Stockinger H., Knapp W.; "Identification of human CD93 as the phagocytic C1q receptor (C1qRp) by expression cloning."; J. Leukoc. Biol. 71:133-140(2002).

RI

[3]

RN

SEQUENCE FROM N.A.

RP

MEDLINE=21638749; PubMed=11780052;

RX

Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Levaeslao M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConachie L.J., McTay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromas A.C., Vaudin M., Wall M., Wallis J.M.,

[illegible]

Db 314 --GATCVLGHGKKNYTRCPQGYQLDSQLDCVVDVDECQSPCAQECVNT 362

RESULT 11

LMA4_HUMAN STANDARD; PRT; 1816 AA.

AC Q16363; Q14735; Q15335; Q9UE18; Q9UJN9;

DT 01-NOV-1997 (Rel. 35, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Laminin alpha-4 chain precursor.

GN LAMA4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal lung;

RX MEDLINE=95300971; PubMed=7781776;

RA Iivanainen A., Sainio K., Sariola H., Tryggvason K.;

RT "Primary structure and expression of a novel human laminin alpha 4 chain.";

RL FEBS Lett. 365:183-188(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=97454279; PubMed=9310354;

RA Richards A.J., Luccarini C., Pope F.M.;

RT "The structural organisation of LAMA4, the gene encoding laminin alpha4.";

RL Eur. J. Biochem. 248:15-23(1997).

RN [3]

RP SEQUENCE OF 236-1816 FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=95048381; PubMed=7959779;

RA Richards A.J., Al-Imara L., Carter N.P., Lloyd J.C., Leversha M.A., Pope F.M.;

RT "Localization of the gene (LAMA4) to chromosome 6q21 and isolation of a partial cDNA encoding a variant laminin A chain.";

RL Genomics 22:237-239(1994).

RN [4]

RP SEQUENCE OF 66-1816 FROM N.A.

RA Tubby B.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.

CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

CC -1- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG, OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE, FETAL. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN BRAIN AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.

CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

CC -1- DOMAIN: DOMAIN G IS GLOBULAR.

CC -1- SIMILARITY: Contains 4 laminin EGF-like domains.

CC -1- SIMILARITY: Contains 5 laminin G-like domains.

CC -1- CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: S78569; AAB34635.1; -

CC EMBL: X91171; CAA62596.1; -

CC EMBL: Y14240; CAA74636.1; -

CC EMBL: X76939; CAA54258.1; -

CC EMBL: Z99289; CAB16553.1; -

CC PIR: S68960; S68960.

CC HSSP: P02488; IKLO.

CC Genew: HGNC:6484; LAMA4.

CC MIM: 600133; -

CC GO: GO:0005605; C:basal lamina; TAS.

CC GO: GO:0005207; F:extracellular matrix glycoprotein; TAS.

CC InterPro: IPR006209; EGF_Like_EGF.

CC InterPro: IPR002049; Laminin_EGF.

CC InterPro: IPR001791; Laminin_G.

CC Pfam: PF00053; laminin_EGF; 3.

CC Pfam: PF00054; laminin_G; 3.

CC SMART: SM00180; EGF_Lam; 3.

CC SMART: SM00282; LamG; 5.

CC PROSITE: PS00022; EGF_1; UNKNOWN_1.

CC PROSITE: PS01248; LAMININ_TYPE_EGF; 3.

CC PROSITE: PS00025; LAM_G_DOMAIN; 5.

CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

CC SIGNAL 1 24 POTENTIAL.

CC CHAIN 25 1816 LAMININ ALPHA-4 CHAIN.

CC DOMAIN 82 131 LAMININ EGF-LIKE 1.

CC DOMAIN 132 186 LAMININ EGF-LIKE 2.

CC DOMAIN 187 240 LAMININ EGF-LIKE 3.

CC DOMAIN 241 255 LAMININ EGF-LIKE 4 (INCOMPLETE).

CC DOMAIN 256 825 DOMAIN II AND I.

CC DOMAIN 826 1028 LAMININ G-LIKE 1.

CC DOMAIN 1040 1220 LAMININ G-LIKE 2.

CC DOMAIN 1227 1395 LAMININ G-LIKE 3.

CC DOMAIN 1462 1633 LAMININ G-LIKE 4.

CC DOMAIN 1640 1813 LAMININ G-LIKE 5.

CC DOMAIN 313 396 COILED COIL (POTENTIAL).

CC DOMAIN 466 521 COILED COIL (POTENTIAL).

CC DOMAIN 574 607 COILED COIL (POTENTIAL).

CC DOMAIN 655 717 COILED COIL (POTENTIAL).

CC DOMAIN 770 799 COILED COIL (POTENTIAL).

CC SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).

CC DISULFID 82 91 BY SIMILARITY.

CC DISULFID 84 98 BY SIMILARITY.

CC DISULFID 101 110 BY SIMILARITY.

CC DISULFID 113 129 BY SIMILARITY.

CC DISULFID 132 146 BY SIMILARITY.

CC DISULFID 134 155 BY SIMILARITY.

CC DISULFID 157 166 BY SIMILARITY.

CC DISULFID 169 184 BY SIMILARITY.

CC DISULFID 187 202 BY SIMILARITY.

CC DISULFID 189 209 BY SIMILARITY.

CC DISULFID 212 221 BY SIMILARITY.

CC DISULFID 224 238 BY SIMILARITY.

CC DISULFID 266 266 INTERCHAIN (PROBABLE).

CC DISULFID 269 269 INTERCHAIN (PROBABLE).

CC CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 308 308 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 458 458 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 524 524 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 574 574 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 639 639 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 735 735 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 751 751 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 754 754 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 780 780 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 798 AA; 88167 MW; 093522509B298531 CRC64;

Query Match 5.8%; Score 92; DB 1; Length 798;
 Best Local Similarity 22.0%; Pred. NO. 2.6;
 Matches 47; Conservative 14; Mismatches 71; Indels 82; Gaps 12;

QY 95 LSGYCD-----CFSSGD-FCNSCSNNLHLEFRKAIKACLDNRNPEAFQPKMGKRL 146
 Db 512 MDAYCRRNSSEISGNDGICGQCVCKKRDNPNEVSG-KYCECDN---FNCDRSNGLI 567
 QY 147 GAAKLHRSKG-CNKRSGCLKNY-----CECYEAKIMCSS-----ICKTACKNKE 191
 Db 568 CG-----CKGICKCRVCEFPNYSGSACDCSDEDTSTCMKNGOICNGRGICDCGRCK--- 619
 QY 192 ESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGGPKLKRNRQAFSCISWEVVEATC-ACL 250
 Db 620 -----CTDPKFGQP-----TCELCQ 634

QY 251 LAQGEAEQHCPSLAEQMILEEFGRLSILH 284
 Db 635 TCVGVCAEHKECVQCAFQK-GEKODVCMEOCMH 667

RESULT 13
 CD93_MOUSE
 ID CD93_MOUSE STANDARD; PRT; 644 AA.
 AC 089103;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Complement component C1q receptor precursor (Complement component 1, q
 DE subcomponent, receptor 1) (ClqR) (ClqR(p)) (Clq/MBL/SPA receptor)
 DE (CD93 antigen) (Cell surface antigen AA4) (Lymphocyte antigen 68).
 GN. ClqR1 OR CD93 OR ClqR OR Ly68 OR AA4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=20528605; PubMed=11074255;
 RA Kim T.S., Park M., Nepomuceno R.R., Palmarini G., Winokur S.,
 RA Cotman C.A., Bengtsson U., Tenner A.J.;
 RT "Characterization of the murine homolog of ClqR(P): identical cellular
 RT expression pattern, chromosomal location and functional activity of
 RL the human and murine ClqR(P).";
 RL Mol. Immunol. 37:377-389(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leukemia;
 RX MEDLINE=99330438; PubMed=10403644;
 RA Petrenko O., Beavis A., Kline M., Kittappa R., Godin I.,
 RA Lemischka I.R.;
 RT "The molecular characterization of the fetal stem cell marker AA4.";
 RL Immunity 10:691-700(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV; TISSUE=Endothelial cells, and Spleen;
 RX MEDLINE=99359842; PubMed=10430565;
 RA Norworthy P.J., Taylor P.R., Walport M.J., Botto M.;
 RT "Cloning of the mouse homolog of the 126-kDa human Clq/MBL/SP-A
 RT receptor, ClqRp.";
 RL Mamm. Genome 10:789-793(1999).
 CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
 CC Clq, mannose-binding lectin (MBL2) and pulmonary surfactant
 CC protein A (SPA). May mediate the enhancement of phagocytosis in
 CC monocytes and macrophages upon interaction with soluble defense
 CC collagens. May play a role in intercellular adhesion. Marker for

CC early multipotent hematopoietic precursor cells. May play a role
 CC in cell-cell interactions during hematopoietic and vascular
 CC development.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in lung, heart and bone marrow.
 CC Expressed at lower level in ovary, whole embryo and fetal liver.
 CC Not detected in brain, adult liver or thymus. Highly expressed in
 CC peritoneal cavity and bone marrow macrophages. Not detected in
 CC epithelial cells.
 CC -!- DEVELOPMENTAL STAGE: First detectable in day 9 embryos, in the
 CC endocardium and vascular endothelium in the anterior part of the
 CC embryo. Expression in endothelial cells, initially restricted to
 CC aorta, omphalomesenteric and umbilical arteries, later extends to
 CC subcardinal veins, intersomitic arteries and perineural vessels.
 CC On day 10, detectable in the entire embryo.
 CC -!- PTM: N- and O-glycosylated (By similarity).
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 5 EGF-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF074856; AAC63274.1; -
 DR EMBL; AF081789; AAC82649.1; -
 DR EMBL; AF099939; AAD47906.1; -
 DR EMBL; AF099938; AAD47906.1; JOINED.
 DR HSSP; P35555; IEMN.
 DR MGD; MGI:106664; Clqrl.
 DR GO; GO:0016023; C:cytoplasmic vesicle; IDA.
 DR GO; GO:0016021; C:integral to membrane; ISS.
 DR GO; GO:0005886; C:plasma membrane; IDA.
 DR GO; GO:0004872; F:receptor activity; ISS.
 DR GO; GO:0016337; P:cell-cell adhesion; ISS.
 DR GO; GO:0042116; P:macrophage activation; ISS.
 DR GO; GO:0006909; P:phagocytosis; ISS.
 DR InterPro; IPR001552; ASX_hydroxyl.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_1; FALSE_NEG.
 DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01187; EGF_CA; 3.
 KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
 FT EGF-like domain; Lectin; Glycoprotein.
 FT SIGNAL 1 22
 FT CHAIN 23 644
 FT DOMAIN 23 572
 FT TRANSMEM 573 593
 FT DOMAIN 594 644
 FT DOMAIN 31 173
 FT DOMAIN 257 298
 FT DOMAIN 299 341
 FT DOMAIN 342 381
 FT DOMAIN 382 423
 FT DOMAIN 424 465
 FT DISULFID 261 272
 FT DISULFID 268 282
 FT DISULFID 284 297
 FT DISULFID 303 314
 FT DISULFID 308 325
 FT DISULFID 327 340
 FT DISULFID 346 355
 FT DISULFID 346 355

```
FT DISULFID 351 364 BY SIMILARITY.
FT DISULFID 366 380 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 393 406 BY SIMILARITY.
FT DISULFID 408 422 BY SIMILARITY.
FT DISULFID 428 440 BY SIMILARITY.
FT DISULFID 436 449 BY SIMILARITY.
FT DISULFID 451 464 BY SIMILARITY.
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 644 AA; 69354 MW; EB4351648BF8635A CRC64;

Query Match 5.7%; Score 91.5; DB 1; Length 644;
Best Local Similarity 24.8%; Pred. No. 2.3;
Matches 37; Conservative 12; Mismatches 51; Indels 49; Gaps 8;

Qy 83 SGPALQPPKTLIS---GYC---DCFSGDFCNCSNNLRHLELRFKAIKACLDNRNPEAF 137
Db 250 SGP-LCVSPKFGCSFNNGCQDCFEFGDGSGFRGC---RPGFRLLDDLVTCASRNPCSS 305
Qy 138 QPKMGKRLGAALKRLHSGKC-----NCKRSGGLKN-----YCE 170
Db 306 NPCTGGGCMCHSVPLSENVTGCRPSGYQLDSSQVHCVDIDECQDSPCAQDCVNTLGSFHC 365
Qy 171 CY-----EAKIMCSICKIACKN 189
Db 366 CWVGYPQSPKKEA---CEDVDECAAAANS 391

RESULT 14
OC90_HUMAN STANDARD; PRT; 493 AA.
AC 002509;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Otoconin 90 precursor (Oc90) (Phospholipase A2 homolog).
GN OC90 OR PLA2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93181166; PubMed=8382789;
RA Feuchter-Murthy A.E., Freeman J.D., Mager D.L.;
RT "Splicing of a human endogenous retrovirus to a novel phospholipase
A2 related gene.";
RL Nucleic Acids Res. 21:135-143(1993).
CC -!- FUNCTION: IT IS UNLIKELY THAT THIS PROTEIN HAS PHOSPHOLIPASE A2
ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. CONTAINS 3
PA2 TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z14310; CAA78662.1; ALT_INIT.
CC HSP: P00593; 4BP2.
CC Genew: HGNC:8100; OC90.
CC MIM: 601658; -.
CC InterPro: IPR001211; PhospholipaseA2.
CC Pfam: PF00068; Phoslip; 2.
CC PRINTS: PR00389; PHPLIPASEA2.
CC ProDom: PD000303; PhospholipaseA2; 2.
CC SMART: SM00085; PA2c; 1.
CC PROSITE; PS00118; PA2_HIS; 2.
```

```
DR PROSITE; PS00119; PA2_ASP; 1.
KW Signal; Glycoprotein; Repeat.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 493 OTOCONIN 90.
FT DOMAIN 76 190 PHOSPHOLIPASE A2-LIKE 1.
FT DOMAIN 321 377 PHOSPHOLIPASE A2-LIKE 2.
FT DOMAIN 389 441 PHOSPHOLIPASE A2-LIKE 3.
FT DISULFID 85 145 BY SIMILARITY.
FT DISULFID 99 190 BY SIMILARITY.
FT DISULFID 101 117 BY SIMILARITY.
FT DISULFID 116 172 BY SIMILARITY.
FT DISULFID 123 165 BY SIMILARITY.
FT DISULFID 132 158 BY SIMILARITY.
FT DISULFID 152 163 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 493 AA; 53338 MW; 59F9AF7C1364ASB7 CRC64;

Query Match 5.7%; Score 91; DB 1; Length 493;
Best Local Similarity 18.3%; Pred. No. 1.9;
Matches 59; Conservative 50; Mismatches 89; Indels 124; Gaps 14;

Qy 2 VICQLKGGAG-MLCIDNCGARELKA-----LHLDPYDDQSSFPQSELPKPTTLVGR 53
Db 150 IICESKDNCHLLCTCDKAAIECLARSSLSNLLDTSFCLAQTPTTIKEDLTLLPR 209
Qy 54 LLPV-PAKLNL-----TQVD-----NGALPSAVNGAA----- 80
Db 210 VVPVEPTDTSLTALSGEVAETADRLITLTKKAGHDQGVGAARATSPGSAEIVATR 269
Qy 81 -----FPSG-----PALQGPPIKTLGYDCDFSSGDFCNCSNNLRHLELRFK 124
Db 270 VTAKIVTLVPAGIKSLGLAVSSVENDEPETERACDRFT---FLHLGSGDNMQVMPQLGE 326
Qy 125 AIRACLDNRNPEATQPKMGKRLGAALKRLHSGKGNCKRSG-----CLKNYCEYE 173
Db 327 MLFCLTSRCPEEPE-----SYGCGYCGEGRPRDLDRLCCLSHHC-CLE 370
Qy 174 -----AKIMCS-----SICKIACKNYEESPERKM 198
Db 371 QVRLGCLLERLPSPVVCVDHTPKCGGQSLCEKLLCACDQTAECMTASFNQSLKSPS 430
Qy 199 LMSTPHYMEPGDFESSHYLSPA 220
Db 431 RLGGPG--QPAACEDSLHPVPA 450

RESULT 15
EZHL_HUMAN STANDARD; PRT; 747 AA.
AC Q92800; Q43287; Q14459;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Enhancer of zeste homolog 1 (ENX-2).
GN EZH1 OR KIAA0388.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97079676; PubMed=8921387;
RA Abel K.J., Brody L.C., Valdes J.M., Erdos M.R., McKinley D.R.,
RA Castilla L.H., Merajver S.D., Couch F.J., Friedman L.S.,
RA Ostermeyer E.A., Lynch E.D., King M.-C., Welsh P.L.,
RA Osborne-Lawrence S., Spillman M., Bowcock A.M., Collins F.S.,
RA Weber B.L.;
RT "Characterization of EZH1, a human homolog of Drosophila Enhancer of
zeste near BRCAL.";
RL Genomics 37:161-171(1996).
RN [2]
```

RP SEQUENCE FROM N.A.
RX MEDLINE-98146265; PubMed-9473645;
RA Ogawa M., Hiraoka Y., Taniguchi K., Aiso S.;
RT "Cloning and expression of a human/mouse Polycomb group gene,
RL ENX-2/Enx-2";
RL Biochim. Biophys. Acta 1395:151-158(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE-97349984; PubMed-9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro";
RL DNA Res. 4:141-150(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE-22389257; PubMed-12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 434-538 FROM N.A.
RX MEDLINE-96039267; PubMed-7490091;
RA Rommens J.M., Durocher F., McArthur J., Tonin P., Leblanc J.F.,
RA Allen T., Samson C., Ferri L., Narod S., Morgan K., Simard J.;
RT "Generation of a transcription map at the HSD17B locus centromeric to
RL BRCA1 at 17q21";
RL Genomics 28:530-542(1995).
CC -!- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF GENE TRANSCRIPTION
CC AND CHROMATIN STRUCTURE.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: BELONGS TO THE EZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U50315; AAC50778.1; -
DR EMBL; AB004818; BAA25019.1; -
DR EMBL; AB002386; BAA20842.1; -
DR EMBL; BC035882; AAI15882.1; -
DR EMBL; L38934; AAB59574.1; -
DR TRANSFAC; T04887; -
DR Genew; HGNC:3526; EZH1.
DR MIM; 601674; -
DR GO; GO:0003682; P:chromatin binding activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR InterPro; IPR001005; Myb_DNA_binding.

DR InterPro; IPR001214; SET.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00717; SANT; 2.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS0280; SET; 1.
KW Transcription regulation; Nuclear protein; DNA-binding.
FT DOMAIN 491 496 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 524 606 CYS-RICH.
FT DOMAIN 612 732 SET.
FT DOMAIN 353 353 P -> S (IN REF. 2).
FT CONFLICT 488 488 N -> Y (IN REF. 1).
FT CONFLICT 532 535 DSTC -> EAL (IN REF. 2).
FT CONFLICT 591 602 ASEHWDCKVSC -> POSTGTARFPV (IN REF. 2).
FT CONFLICT 700 747 VVMVNGDHRIQIFAKRAIQAGEELFFDYRYSQADALKYVGI
FT ERETDLV -> GESQ (IN REF. 2).
SQ SEQUENCE 747 AA; 85270 MW; 7CFC52269CDA011B CRC64;
Query Match 5.7%; Score 91; DB 1; Length 747;
Best Local Similarity 21.0%; Pred. NO..3;
Matches 37; Conservative 17; Mismatches 44; Indels 78; Gaps 9;
QY 99 CDCFSSGDFONS-CSCNNLRHELEFRFAKACLDENPEAFQPKMGKGRGAALKRHSKGC 157
DB 535 CPCIMTQNFCEKFCOCN-----PDCQNRFP-----GC 561
QY 158 NCKRSGCLKNYCEYEAKIMCSS-----ICKCIAKNYE-ESPERKMLMSTPH 204
DB 562 RCK-TQCNTKQPCYLAVERCDPDCLCTCGASEHWDCKVSCKNCSIQRLKKHLL----- 616
QY 205 YMEPGDFESSHYLSPAKFSG-----PPKLNRKROAFSCISWEVVEATCACLAGE 255
DB 617 -----LAPSDVAGMGTFKESVQKN-----EFISEYCGELISQDE 651
Search completed: July 24, 2003, 13:37:34
Job time : 17.899 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:34:27 ; Search time 25.3283 Seconds
(without alignments)
1120.082 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	396	24.8	429	T23152	hypothetical prote
2	363	22.7	571	B84585	hypothetical prote
3	357	22.3	603	T08955	hypothetical prote
4	232	14.5	658	F71410	hypothetical prote
5	115	7.2	1737	T00209	MEGF8 protein - hu
6	100	6.3	775	T21436	hypothetical prote
7	97	6.1	497	T27827	hypothetical prote
8	96.5	6.0	1790	1 MMRFB1	laminin beta-1 cha
9	96	6.0	769	5 S5554	male-specific leth
10	96	6.0	3635	2 T10053	laminin alpha 5 ch
11	95	5.9	1096	1 S61917	protein kinase C (
12	95	5.9	1428	2 T08852	lustrin A - Califo
13	94	5.9	716	2 T26998	hypothetical prote
14	94	5.9	1819	2 A71928	cag island protein
15	93	5.8	1816	1 S68960	laminin alpha-4 ch
16	92	5.8	788	2 A28193	integrin beta-1 ch
17	91.5	5.7	383	2 D88633	protein F56B3.2 [1
18	91	5.7	1927	2 G64585	cag pathogenicity
19	91	5.7	2907	2 A57278	fibrillin-2 precur
20	90.5	5.7	735	2 G02937	fertilin beta - cr
21	89	5.6	574	2 B88465	protein B0244.8 [i
22	89	5.6	1786	1 MMRB1	laminin beta-1 cha
23	88	5.5	625	2 F84706	probable vacuolar
24	88	5.5	798	2 B28193	integrin beta-1* c
25	88	5.5	1786	1 MMRB1	laminin beta-1 cha
26	87	5.4	290	2 G72858	AcOrf-70 protein -
27	87	5.4	501	2 S56163	tumor necrosis fac
28	87	5.4	652	2 T38704	glycyl tRNA synth
29	87	5.4	902	2 T01127	curly leaf protein

30	86.5	5.4	4545	1 S25111	alpha-2-macroglobu
31	86	5.4	152	2 T18975	hypothetical prote
32	86	5.4	610	2 A35046	E-selectin prote
33	86	5.4	830	2 A30359	P-selectin precurs
34	86	5.4	1034	2 JC5598	mucin - rat
35	85.5	5.3	769	1 IJHULM	leukocyte adhesio
36	85.5	5.3	772	2 S32659	integrin beta 2 ch
37	85.5	5.3	806	2 A46271	integrin beta-7 ch
38	85.5	5.3	1599	2 T16210	hypothetical prote
39	85	5.3	398	2 A35281	integumentary muc
40	85	5.3	2543	2 T31687	surface antigen - P
41	84.5	5.3	169	1 SI8946	ultra high-sulfur
42	84.5	5.3	307	2 T09923	cytidine deaminase
43	84.5	5.3	1639	1 MMFB2	laminin gamma-1 ch
44	84	5.3	769	1 JC1121	leukocyte adhesio
45	84	5.3	788	2 I51530	integrin beta-3 su

ALIGNMENTS

RESULT 1

T23152

hypothetical protein JC8.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23152

R:Lightning, J.

Submitted to the EMBL Data Library, November 1996

A:Reference number: Z19699

A:Accession: T23152

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-429 <WIL>

A:Cross-references: EMBL:Z82274; PIDN:CAB05228.1; GSPDB:GN00022; CESP:JC8.6b

A:Experimental source: clone JC8

C:Genetics:

A:Gene: CESP:JC8.6b

A:Map position: 4

A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Query Match 24.8%; Score 396; DB 2: Length 429;

Best Local Similarity 38.9%; Pred. No. 1.3e-25;

Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

Qy 98 YCDCFSGGDFCNSCSC-----NLRHLELEREKAICACLDNRPEAFQPMKGRGCAAKLR- 152

Db 182 YCDCFANGFCRDCNCKDCHNIEYDQSKAIRQSLERNPNAPKPKIGIARGITDIER 241

Qy 153 -HSKGCNCRSGCLKNVCEYEAQIMCSTICKCTACKN-----YEES-----PERKML 199

Db 242 LHQKCHCKSGCLKNVCEYEAQVPTDRCKCKGCONCTYRMYTRKNSGGAVSNTNAL 301

Qy 200 MS-----TPHYMPGDFESSHY---LSPAKFSGPPKLRKNRQAFSCISW-----EVV 243

Db 302 MSLTNASSIATPDP-SGPGSVVTDEHGDDYDMLLSHKPKVEMDPRP---PWYYMTDEVV 357

Qy 244 EATCACLQAQGEA-----EQBHCSPSLAEQMTLEEFGRCLSQL 283

Db 358 EAATMCVQAQAEALNVEKVQTEDEKLINMEKLVLRFGRLCLQMI 403

RESULT 2

B84585

hypothetical protein At2g20110 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: B84585

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84585

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-571 <STO>

A:Cross-references: GB:AE002093; NID:g4580462; PIDN:AAD24386.1; GSPDB:GN00139

C:Genetics:

A:Gene: AT2G20110

A:Map position: 2

Query Match 22.7%; Score 363; DB 2; Length 571;
Best Local Similarity 32.4%; Pred. No. 1e-22;
Matches 81; Conservative 39; Mismatches 76; Indels 54; Gaps 8;

QY 36 SSFPOSELPKPMPTTLVGRLL-----PVPKLNLTITQVD-NGALPSAVNGAAF 81
DB 46 SSMASASPPIVTRPIISQAPPTVATPIPPPPQSGIILHVPRIHRPESNSMPR 105
QY 82 PSGPALQGPCK-----ITLSGYCDGFCFSGDFCNCSGSC-----NNLRHELEFRFAIK 127
DB 106 PAGETRDGTPQKKKOCNKHRSCLKLYCECFASGYCDGCVNCFNVNENEPARROAVE 165
QY 128 ACLDRNPAFQPKMGKGRIGRAK-----LRHSGKCNKRSGLKNYCECYEAKI 176
DB 166 STLERNPAFRPKIAASPHGDRNREYGVVYMLARHNKGCHCKSGGLKLYCECFQANI 225
QY 177 MCSSTCKCIACKNYEESPCKMLMSTPH-----YME-----PGDFESSHYLVSPAKFSGP 225
DB 226 LCSENCKLDCKNFCGSEVROSLFGEHSHNLAYLQHANAAITGAIGSSGFAS-----APP 281
QY 226 PKLRNKNQAF 235
DB 282 PKRRKGQEIF 291

RESULT 3

T08955

hypothetical protein F19B15.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999

C:Accession: T08955

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,

submitted to the Protein Sequence Database, May 1999

A:Reference number: 216519

A:Accession: T08955

A:Molecule type: DNA

A:Residues: 1-603 <BEV>

A:Cross-references: EMBL:AL078470; GSPDB:GN00062; ATSP:F19B15.30

A:Experimental source: cultivar Columbia; BAC clone F19B15

C:Genetics:

A:Gene: ATSP:F19B15.30

A:Map position: 4

A:Introns: 107/2; 144/2; 203/3; 321/3; 356/2; 388/1; 522/1

Query Match 22.3%; Score 357; DB 2; Length 603;
Best Local Similarity 35.2%; Pred. No. 3.4e-22;
Matches 88; Conservative 38; Mismatches 66; Indels 58; Gaps 12;

QY 36 SSFPOSELPKPMPTTLVGRLLPVPKLNLTITQVDNGALPSAVNGA---AFSPGALQGP- 90
DB 60 SVPSTIRP-GMTIAIGQVTVRPTLPMTATMSNPSPSQIIVNAPIRHIPESPARGPR 118
QY 91 PKI-----TLSGYCDGFCFSGDFCNCSGSC-----NNLRHELEFRFAIK 127
DB 119 PNVEGRDGTPOKKKOCNKHRSCLKLYCECFASGYCDGCVNCFNVNENEPARREAVE 178
QY 128 ACLDRNPAFQPKMGK-----GR-----LGAALK--RHSKGCNKRSGGLKNYCECYEAKI 176
DB 179 ATLERNPAFRPKIAASSPHGGRKREDIGEVVLLGKNKGCHCKSGGLKLYCECFQANI 238
QY 177 MCSSTCKCIACKNYEESPCKMLMSTPH-----YME-----PGDFESSHYLV-SPAKFS 223
DB 239 LCSENCKLDCKNFCGSEVROSLFGEHSHNLAYLQHANAAITGAIGSSGFASPA--- 295

QY 224 GPPKLRKNRQ 233

DB 296 --PKRRKGQE 303

RESULT 4

F71410

hypothetical protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

A:Variety: Columbia

C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998

C:Accession: F71410

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; G

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B

Nature 391, 483-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen

erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;

C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t

A:Reference number: A71400; MUID:98121113; PMID:9461215

A:Accession: F71410

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-658 <BEV>

A:Cross-references: GB:297337; NID:g22444829; PID:e326824; PID:g2244834

C:Genetics:

A:Map position: 4COP9-4G3845

Query Match 14.5%; Score 232; DB 2; Length 658;
Best Local Similarity 39.5%; Pred. No. 1.1e-11;
Matches 51; Conservative 11; Mismatches 43; Indels 24; Gaps 5;

QY 82 PSGPALQ-----GPPKITLSGYCDGFCFSGDGC-NSCSC-----NNLRHELEFRFAIKACLD 131
DB 354 PVEPALQELNLSSPKK---KSYCEFAAGVYCIIEPCSCIDCFNKPDIHEDVVLATRRQIES 410
QY 132 RNPAFAQPKMGKGRIGA-----AKLRHSGKCNKRSGGLKNYCECYEAKIMGSS 180
DB 411 ENPLAFAPKVRNSDSVQETGDDASKTPASARHKGCCNCKLKKYCECYGGVGCSI 470
QY 181 ICKCIACKN 189
DB 471 NCRCEGCKN 479

RESULT 5

T00209

MEGF8 protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00209

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like moti

A:Reference number: Z14126; MUID:98360089; PMID:9693030

A:Accession: T00209

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1737 <NAK>

A:Cross-references: EMBL:AB011541; NID:g3449307; PIDN:BAA32469.1; PID:g3449308

A:Experimental source: brain; clone HG1392

C:Genetics:

A:Gene: MEGF8

A:Map position: 19q12

Query Match 7.2%; Score 115; DB 2; Length 1737;

Best Local Similarity 23.1%; Pred. No. 0.2;

Matches 73; Conservative 27; Mismatches 124; Indels 92; Gaps 18;

QY 17 NCGARELKALHLLPOYDQSSFPQSELPKPMPTTLVGRLLPVPKLNLTITQVDNGALPSAV 76

```
Db 711 NCNA-----WLLPDLTRSA-----VGPPMEESVAHAAYVGSRLYISGGFGGVALGLR 759
QY 77 NGAAPSPGA--LOGPPKITLSGYCD-CFS-----SGD-----FCNSCSCNNL-----RHELER 122
Db 760 LALTLPDCRLLSPEACNQSGACTWUGACLSGDQAHRLCGGSPSPMPRSPPEECRR 819
QY 123 FKAIAKCLDRNPEAFQPKMGKRLGAALRHRSKGC-NCKRSQCL-----KNYC-----169
Db 820 LRTCSECLARHPTLQPGDGE-----ASTPRCKWCTNCPEGACIGRNGSCTSENDCRINO 874
QY 170 -----ECYEAKI-----MCSSICKCIACKNYESPPE-RKMLMSTPHY-----205
Db 875 REVFMAGNCEAAGCAACDEQCETREGKCMWTFQFRTGETRILSVQPTDYDTCFSHSL 934
QY 206 -MEPGDFESSHYLSPAKFGPPKLRNRQAFSCISWEVVEATCACLLAOGAEAEODHPCSP 264
Db 935 NVSPMPVSESS---PPLPCTTPCHLLPN-----CTS-----CLDSKGADGGWQHCVW 977
QY 265 SLAEQMILEEFGRCILS 280
Db 978 SSSLO-----QCLS 986

RESULT 6
T21436
hypotheical protein R06A4.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21436; T23953
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21436
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-775 <WIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB04199.1; GSPDB:GN00020; CESP:R06A4.7
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19822
A:Accession: T23953
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-775 <W12>
A:Cross-references: EMBL:Z83120; PIDN:CAB05589.1; GSPDB:GN00020; CESP:R06A4.7
A:Experimental source: clone R06A4
C:Genetics:
A:Gene: CESP:R06A4.7
A:Map position: 2
A:Introns: 223/1; 414/3; 464/2; 537/1; 601/2; 655/1; 679/1; 749/2

Query Match 6.3%; Score 100; DB 2; Length 775;
Best Local Similarity 21.8%; Pred. No. 1.6;
Matches 32; Conservative 14; Mismatches 41; Indels 60; Gaps 7;

QY 89 GPPKITLSGYCDGFCSSGDFCNSCNNLRHELERFKAIAKCLDRNPEAFQPKMGKRLGA 148
Db 537 GPCNATAEN-CACRENGVCSYCKCD-----INCSQFP-----569
QY 149 AKLRHSGCNCKRSGLKNVCEYEAIKMCS-----SICKC-----IACKNYEESPERK 197
Db 570 -----GCNCAAGQCYTACQYRANWECNPMWTCNCKCDAIDSLIKRNFEGMT---R 619
QY 198 MLMSTPHYMEPGDFESSHYLSPAKFSG 224
Db 620 MI-----QKRTTCGSPSKIAG 634

RESULT 7
T27827
hypotheical protein ZK287.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
```

```
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27827
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z20425
A:Accession: T27827
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-497 <WIL>
A:Cross-references: EMBL:Z70757; PIDN:CAA94805.1; GSPDB:GN00023; CESP:ZK287.1
A:Experimental source: clone ZK287
C:Genetics:
A:Gene: CESP:ZK287.1
A:Map position: 5
A:Introns: 20/1; 44/2; 99/3; 339/3; 443/2

Query Match 6.1%; Score 97; DB 2; Length 497;
Best Local Similarity 26.6%; Pred. No. 1.8;
Matches 37; Conservative 15; Mismatches 51; Indels 36; Gaps 10;

QY 82 PSGPALQPPPKITLSGYCDGFCSSGDFCNS-CSCNNLRHELERFKAIAKCLDRNPEAFQPK 140
Db 63 PAKPAKNACAK---SERCDLRIKEGCKLCAKQI-----C--KNKEA--PK 103
QY 141 MGKGRGLGAAL-RHSGKCNCK--RSGLKNVCEYEAIKMCSICKICIA-CKNYEESPER 196
Db 104 K-----LAKVAKPTSGCQCAKGGKQCVKKECACATVYGFCSSASCCKGGDCTNGASK---154
QY 197 KMLMSTPHYMEPGDFESSH 215
Db 155 ---FSVPKHVQNCFLRHHK 170

RESULT 8
MMFFBI
laminin beta-1 chain precursor - fruit fly (Drosophila melanogaster)
N:Alternate names: laminin chain B1
C:Species: Drosophila melanogaster
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C:Accession: A28783; S14462; E28783
R:Montell, D.J.; Goodman, C.S.
Cell 53, 463-473, 1988
A:Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals
A:Reference number: A28783; MUID:88210471; PMID:3365769
A:Accession: A28783
A:Molecule type: mRNA
A:Residues: 1-1790 <MON1>
A:Cross-references: EMBL:M19525
R:Montell, D.J.; Goodman, C.S.
submitted to the EMBL Data Library, June 1988
A:Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain r
A:Reference number: S14462
A:Accession: S14462
A:Molecule type: mRNA
A:Residues: 1-667, 'L', 669-725, 'VT', 728-947, 950-1790 <MON2>
A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
C:Genetics:
A:Gene: lamB1
A:Cross-references: FlyBase:FBgn0002527
A:Map position: 2L 28D
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
C:Function:
A:Description: interact with cells and with other basement membrane proteins to promo
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1790/Product: laminin beta-1 chain #status predicted <NAT>
F:27-288/Domain: VI <DOM6>
F:289-561/Domain: V <DOM5>
F:290-354/Domain: laminin-type EGF-like homology <LEO1>
F:357-417/Domain: laminin-type EGF-like homology <LEO2>
F:420-477/Domain: laminin-type EGF-like homology <LEO3>
F:480-528/Domain: laminin-type EGF-like homology <LEO4>
```

F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:562-789/Domain: IV <DOM4>
F:643-645/Region: cell attachment (R-G-D) motif
F:790-1189/Domain: III <DOM3>
F:791-836/Domain: laminin-type EGF-like homology <LE06>
F:839-882/Domain: laminin-type EGF-like homology <LE07>
F:885-932/Domain: laminin-type EGF-like homology <LE08>
F:935-990/Domain: laminin-type EGF-like homology <LE09>
F:968-972/Region: cell adhesion #status predicted
F:993-1042/Domain: laminin-type EGF-like homology <LE10>
F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
F:1190-1407/Domain: II <DOM2>
F:1408-1434/Domain: alpha <ALP>
F:1435-1790/Domain: I <DOM1>
F:51-56/Disulfide bonds: #status predicted
F:140,203,234,489,593,1053,1248,1303,1332,1343,1475,1495,1517,1583,1646,1705/Binding site
F:1191,1194,1788/disulfide bonds: interchain #status predicted

Query Match 6.0%; Score 96.5; DB 1; Length 1790;
Best Local Similarity 21.0%; Pred. No. 7.3;
Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;

Qy 94 TSLGYDCFFSGDF-----CNSCS-----NHLRHLERFKAIKACLDNRNPEAFQPKM 141
Db 1021 TTGDHCELCKDGFQDALQNCQCECDLGTNTNTIAHCDRTGQCPL-----PNV 1072

Qy 142 KGRLGAARLRSK-----GCKRSGCLKNYCEYEAKEIMCSSICKIACK-----N 189
Db 1073 QGVKCDQCAENHWKIATSGEGESCNCDPDIGALHEQCNSYTGCCQCKPGFGGRACNQCQAH 1132

Qy 190 YEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWEVVEATCAC 249
Db 1133 YWGNPNEKQPCQCDQFGAADFQC-----DRTGNCVCHEGGYKCN 1175

Qy 250 LLAQGEAEQHCSP 264
Db 1176 ECARGYIGQFPHCSP 1190

RESULT 9
S55554
male-specific lethal 2 protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S55554
R:Zhou, S.; Yang, Y.; Scott, M.J.; Pannuti, A.; Fehr, K.C.; Eisen, A.; Koonin, E.V.; Foulk
EMBO J. 14, 2884-2895, 1995
A:Title: Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex
A:Reference number: S55554; MUID:95317307; PMID:7796814
A:Accession: S55554
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-769 <ZHO>
A:Cross-references: EMBL:248443; NID:9872110; PIDN:CAA88358.1; PID:9872111
C:Genetics:
A:Gene: FlyBase:msl-2
A:Cross-references: FlyBase:FBgn0005616
A:Introns: 641/1

Query Match 6.0%; Score 96; DB 2; Length 769;
Best Local Similarity 25.7%; Pred. No. 3.4;
Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;

Qy 121 ERFKAIRAC---LDRNPEAFQPKMGKRLCAAKLRHSKG-----CNCKRSG 163
Db 471 KRTRLKAQAQAEIPVSEVTKVQSGK-GA--LRRIRGDKKEKVPKPKRCRGISG 527

Qy 164 -----CLKNYCEYEAKEIMCSSICKIACKN-----YEES-----PER 196
Db 528 SSNTLTTCNSRCFCYKSYNSCAG-CHVCCKNPKHEDYVESDDDDLEDFEMPKDPEP 586

Qy 197 KMLMSTPHYMEPGDFESSHYLSPAKFSGPPKL 228
Db 587 MTQSEEPVVAEPRQEENS--MAPDSSAPISL 616

RESULT 10
TI0053
laminin alpha 5 chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
C:Accession: TI0053
R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z16923
A:Accession: TI0053
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3635 <MIN>
A:Cross-references: EMBL:U37501; NID:g2599231; PID:g2599232
C:Genetics:
A:Gene: Lama5
C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like h
C:Keywords: basement membrane; cell binding; extracellular matrix
F:1888-1939/Domain: laminin-type EGF-like homology <LEG>
F:1942-1970/Domain: EGF homology <EGF>

Query Match 6.0%; Score 96; DB 2; Length 3635;
Best Local Similarity 18.0%; Pred. No. 17;
Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;

Qy 14 CIDNCGARELKALHLLPYDDQSSFPQSELPKPMTLVGRLLPVPKLNLTQVNDGALP 73
Db 117 CLERFGPRTLR---ITQDDVI-----CTTEYSRIVPL-ENGEIVVSLVNGR-P 161

Qy 74 SAVNGAAPP-----SGPALGP-----PKITLSGY 98
Db 162 GALNFSYSPLLRDTKATNIRLRLNTLGLHLMGKALRDPVTRRRYYYSIKDISIGR 221

Qy 99 CDCFSSGDFC-----NSC-----SC 113
Db 222 CVCHGADVCDKDPDLPFLQCAQHNTCCGSDCCPGFNQOPKWPATTDSANECQSC 281

Qy 114 NNLRH-----ELERFKAIKACLDNRNPEAFQPKMGKRLCAAKL---RISKGCNCRS 162
Db 282 NCHGHAYDCYYDPEVDR-----RNASONQNVYOG--GGVCLDCQHHTTGINCER- 329

Qy 163 GCLKNY-----CEYEAKI---MCSSTI---CK-----CIAC-KNY 190
Db 330 -CLGFFRAPDQPLDSPHVCRPCDC-ESDFTDGTCDLTGRCYCRPNFTGELCAACAGY 387

Qy 191 EESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWEVVEATCAC 250
Db 388 TDFPHCYPLPSFPHN-----DTREQVLP-----GQIVNCDCA 422

Qy 251 LAQGEAEQ-----HC-----SPSLAEMILEEGR 277
Db 423 GTQGNACKRDPRLGRVCVKPNFRGAHCELCAPFGHPSCHPCQCSSPGVANSCLDPESGQ 482

Qy 278 CL 279
Db 483 CM 484

RESULT 11
S61917
protein kinase C (EC 2.7.1.1) PKCA - Aspergillus niger
C:Species: Aspergillus niger
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S61917
R:Morawetz, R.; Lendenfeld, T.; Mischak, H.; Muehlbauer, M.; Gruber, F.; Goodnight, J
Mol. Gen. Genet. 250, 17-28, 1996
A:Title: Cloning and characterisation of genes (pkci and pkca) encoding protein kinas
A:Reference number: S61917; MUID:96158841; PMID:8569684

Query Match 5.98; Score 94; DB 2; Length 1819;
 Best Local Similarity 21.4%; Pred. No. 12;
 Matches 63; Conservative 39; Mismatches 113; Indels 80; Gaps 12;

QY 22 ELKALHLLPOYDDOSSFPQSELKPK-----MTLVGLLLPVPAP-----LNLITQVNG 70
 Db 568 DILAKESLKYKACASQAKTEAEKKEKLLTPPAKLLLEEAKEKSVKAYLDCVSOAKTE 627
 QY 71 ALPSAVNGAAPPSPALQGPPTITLSGYCDGFCSCNCLNRLHLEKFAKACL 130
 Db 628 AEKKEKLLTPPAKLLLEEAKEKSVRAYLDCVSKA-----KNEAEKKEKLL- 675
 QY 131 DRNPEAFQPKMGKRLGAALKRHSKGCN-----KRSGLKNYCEYEAKIMCSSIC 182
 Db 676 --TPE-----AKLLENAQLDCLNNAKTDEERKEKLDLPKDLQKKVLAKESV 721

QY 183 K-----CIA-CNKYEESPERKMLSTPHYMEPGDFESSHYLSPAKFSPPKLRNROAFSC 237
 Db 722 RYVLDVSKAKNEAEKKEKLL-TP-----EARKLLEEAKEK----- 758

QY 238 ISWEVVEATCACLIAOGEAEAEHCHSPSLAEQ--MILEEFGRLCSQILHIEFKSK 290
 Db 759 -----VRAYKDCVLRARNEKEKQCEKLLTPPAKLLLEESKSVKAYLDCVSKAK 808

RESULT 15
 S68960
 laminin alpha-4 chain precursor - human
 N:Alternate names: laminin A4
 C:Species: Homo sapiens (man)
 C:Date: 21-Aug-1998 #sequence,revision 21-Aug-1998 #text_change 16-Jun-2000
 C:Accession: S68960; S65926; S49149; S40150; I53516
 R:Richards, A.; Al-Imara, L.; Pope, F.M.
 Eur. J. Biochem. 238, 813-821, 1996

A:Title: The complete cDNA sequence of laminin alpha-4 and its relationship to the other
 A:Reference number: S68960; MUID:96300249; PMID:8706685
 A:Accession: S68960
 A:Molecule type: mRNA
 A:Residues: 1-1816 <RIC>
 A:Cross-references: EMBL:X91171; NID:g1212962; PIDN:CAA62596.1; PID:g1212963
 A:Experimental source: tissue type heart
 R:Rivannainen, A.; Sainio, K.; Sariola, H.; Tryggvason, K.
 FEBS Lett. 365, 183-188, 1995

A:Title: Primary structure and expression of a novel human laminin alpha-4 chain.
 A:Reference number: I53516; MUID:95300971; PMID:7781776
 A:Accession: S65926
 A:Molecule type: mRNA
 A:Residues: 1-142, 'P', 144-177, 'F', 179-490, 'Y', 492-1056, 'P', 1058-1816 <IIV>
 A:Cross-references: EMBL:S78569; NID:g1042081; PIDN:AAB34635.1; PID:g1042082
 R:Richards, A.J.; Al-Imara, L.; Carter, N.; Leversha, M.; Lloyd, J.C.; Pope, F.M.
 submitted to the EMBL Data Library, December 1993

A:Description: Localisation of the gene (LAMA4) to chromosome 6q21 and isolation of a pa
 A:Reference number: S49149
 A:Accession: S49149
 A:Molecule type: mRNA
 A:Residues: 236-1816 <RI2>
 A:Cross-references: EMBL:X76939; NID:g509805; PIDN:CAA54258.1; PID:g509806
 R:Richards, A.J.; Al-Imara, L.; Carter, N.; Lloyd, J.C.; Pope, F.M.
 submitted to the EMBL Data Library, February 1993

A:Description: Isolation of a partial cDNA encoding a protein homologous to laminin A. A
 A:Reference number: S40150
 A:Accession: S40150
 A:Molecule type: mRNA
 A:Residues: 1403-1541, 'S', 1543-1816 <RI3>
 A:Cross-references: EMBL:X70904; NID:g437804; PIDN:CAA50261.1; PID:g437805
 C:Genetics:
 A:Gene: GDB:LAMA4; LAMA3
 A:Cross-references: GDB:203904; OMIM:600133
 A:Map position: 6q21-6q21
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
 C:Function:
 A:Description: Interact with cells and with other basement membrane proteins to promote
 C:Superfamily: laminin alpha-4 chain; laminin G repeat homology; laminin-type EGF-like H

C:Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glyco
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-1816/Product: laminin alpha-4 chain #status predicted <MAT>
 F;82-129/Domain: laminin-type EGF-like homology <LE1>
 F;132-184/Domain: laminin-type EGF-like homology <LE2>
 F;187-238/Domain: laminin-type EGF-like homology <LE3>
 F;241-265/Domain: laminin-type EGF-like homology #status atypical <LE4>
 F;717-719/Region: cell attachment (R-G-D) motif
 F;862-1031/Domain: laminin G repeat homology <LG1>
 F;1068-1223/Domain: laminin G repeat homology <LG2>
 F;1252-1398/Domain: laminin G repeat homology <LG3>
 F;1488-1636/Domain: laminin G repeat homology <LG4>
 F;1565-1816/Domain: laminin G repeat homology <LG5>
 F;104,215,308,458,524,550,571,574,631,639,735,751,754,780,803,1086,1281,1359,1411/Bin
 F;266,269/Disulfide bonds: interchain #status predicted

Query Match 5.88; Score 93; DB 1; Length 1816;
 Best Local Similarity 23.8%; Pred. No. 15;
 Matches 74; Conservative 32; Mismatches 143; Indels 62; Gaps 18;

QY 2 VICOLKGGQMLCIDNCG-----ARELKALH---LLPOY--DDOSSFPQSELKPKMTTLV 51
 Db 80 VPDCNGNSNE-CLDGSYCVHCRNTTGBCEKCLDGYIGDSIRGAPQFCQPCP----- 133

QY 52 GRLLPVPKLNLTQV--DNGALPSAVN-GAAPFS-----GPALQGPPIITLSGYCDGFCSS 104
 Db 134 ---CPLPLANFAESCYRKNKNGAVRCICNENYAGNCERCAPGYGNPLLI----- 180

QY 105 GDFCNSCSN-----NLRHE--LERFKAIKACLDNRNEAFQ-PKMGKRLGAALK-RHSK 155
 Db 181 GSTCKKCDKSGNSDPNLIIFEDCDCEVTGQCRNCL-RNTTGFKERCAPGYGVDARIAKNC 239

QY 156 GCNCKRSGLKNYCEYEAKIMCSSIC-KCI-----ACKNYEESPERKMLMST--P 203
 Db 240 VCNCGGCPDSVTGCECLEEGFEPTGCDKCVMDLTDLLRLAALSIEGKSGVLSVSSGAA 299

QY 204 HYMEPGDFESSHYLSPAKFSPPKLRNROAFSCISWEVVEATCACLIAOGEAEAEHCHS 263
 Db 300 AHRHVNEINATIVLLKTKLSE---RENOVALRKIQINNAENTMTKSLSDVEELVERENQ 355

QY 264 PSIAEQMILEE 274
 Db 356 ASRKQLVQKE 366

Search completed: July 24, 2003, 13:48:39
 Job time : 27.3283 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:37:08 ; Search time 19.3687 Seconds
(without alignments)
1808.803 Million cell_updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICOLKGAQMLCIDNCGA.....GRCLSQILHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	396	24.8	438	9	US-09-220-091-13
2	322	20.1	147	11	US-09-764-891-3988
3	248	15.5	53	15	US-10-106-6398-6932
4	115	7.2	1192	15	US-10-189-971-18
5	115	7.2	1207	15	US-10-189-971-20
6	115	7.2	1477	15	US-10-189-971-8
7	115	7.2	1535	15	US-10-189-971-14
8	115	7.2	1593	15	US-10-189-971-4
9	107.5	6.7	1057	15	US-10-189-971-6
10	107.5	6.7	1251	15	US-10-189-971-16
11	107.5	6.7	1342	15	US-10-189-971-24
12	107.5	6.7	1512	15	US-10-189-971-10
13	107.5	6.7	1570	15	US-10-189-971-12
14	107.5	6.7	1628	15	US-10-189-971-2
15	100	6.3	759	15	US-10-189-971-22

16	100	6.3	773	15	US-10-231-778-223	Sequence 223, Appl
17	97.5	6.1	652	9	US-09-789-919-96	Sequence 96, Appl
18	96	6.0	3635	10	US-09-845-583-2	Sequence 2, Appl
19	96	6.0	3635	15	US-10-037-182-4	Sequence 4, Appl
20	93	5.8	1816	15	US-10-299-058-2	Sequence 2, Appl
21	93	5.8	1816	15	US-10-299-058-4	Sequence 4, Appl
22	93	5.8	4123	15	US-10-213-509-5	Sequence 5, Appl
23	91.5	5.7	644	9	US-09-789-919-62	Sequence 62, Appl
24	91.5	5.7	644	9	US-09-789-919-73	Sequence 73, Appl
25	90.5	5.7	167	11	US-09-898-659-28	Sequence 28, Appl
26	89.5	5.6	347	9	US-09-789-561-184	Sequence 184, Appl
27	89.5	5.6	361	15	US-10-091-458-34	Sequence 34, Appl
28	89.5	5.6	459	15	US-10-153-668-86	Sequence 86, Appl
29	89	5.6	1725	15	US-10-037-182-12	Sequence 12, Appl
30	89	5.6	1786	10	US-09-938-275-7	Sequence 7, Appl
31	89	5.6	1786	15	US-10-037-182-10	Sequence 10, Appl
32	88.5	5.5	260	10	US-09-858-546-5	Sequence 5, Appl
33	88.5	5.5	530	10	US-09-858-546-2	Sequence 2, Appl
34	88	5.5	1349	15	US-10-077-583-6	Sequence 6, Appl
35	88	5.5	1551	9	US-09-970-318-4	Sequence 4, Appl
36	88	5.5	1765	15	US-10-037-182-8	Sequence 8, Appl
37	88	5.5	1786	9	US-09-873-676-113	Sequence 113, Appl
38	88	5.5	1786	10	US-09-938-275-6	Sequence 6, Appl
39	88	5.5	1786	15	US-10-037-182-6	Sequence 6, Appl
40	87.5	5.5	610	11	US-09-802-640-36	Sequence 36, Appl
41	87.5	5.5	610	15	US-10-205-823-357	Sequence 357, Appl
42	87	5.4	867	10	US-09-895-913A-358	Sequence 358, Appl
43	86.5	5.4	4545	9	US-09-873-403-2	Sequence 2, Appl
44	86	5.4	418	15	US-10-201-386-13	Sequence 13, Appl
45	86	5.4	732	11	US-09-977-418-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Query Match	24.8%	Score 396;	DB 9;	Length 438;
Best Local Similarity	38.9%;	Pred. No. 8.3e-31;		
Matches	88;	Conservative 32;	Mismatches 62;	Indels 44; Gaps 10;
QY	98	YDCFSGDFCNCS	-----NNLRHELEPKAKIACLDNRPEAFQPKMGKRGAKLR-	152
DB	191	YDCFFANGFDCNCKDCNHNIEYDSQSRKALQSLERNPNAPKPKIGIARGIDIER	250	
QY	153	-HSKGCNCRSGCLKNYCEYEA	IMCSSTCKIACKN-----YEE	199
DB	251	LHQKGCCHKSGCLKNYCEYEA	KVPCTDRCKCKGQNTYTRYKNSGGAVSNAL	310
QY	200	MS-----TPHYMEPGDFESSHY	---LSPAKFSGPKLKRNRQAFSCISW-----EVV	243

— — — — —

```
RESULT 8
US-10-189-971-4
; Sequence 4, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-4

Query Match      7.2%; Score 115; DB 15; Length 1593;
Best Local Similarity 21.1%; Pred. No. 0.034;
Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

Qy      2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMWTLVGR 53
Db      753 VSCPEKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPREPNCNLCTCLGG 808
Qy      54 LLPVPAKLNLITQVDNGALPSAVNGAAPPSPALQGGPKITLSGYC-----DC-----F 102
Db      809 F-----VTCGRRCPEPGCSHPLP-----SCHCCPTCGCRYHGVTT 846
Qy      103 SSGD-----FCNSCSNNLRHELFRKAICACLDNRNPEAFQPKMGK-----GRL 146
Db      847 ASGETLPDPLDPTCSLCTCQGREHQ-----DGEFEGPAGSCWCRCQAGQV 893
Qy      147 GAAKLR-----HSKGCNCR-SGCLK-----NYCEYE----- 173
Db      894 SCVLQCPPLPKLQVTERGCCPRGCLAHGEHPEGSGRWVPPDSACSCVCHGVVT 953
Qy      174 -AKIMCSSICK-----CIACKNVEESPCKMLMSTHYMEPGD-FESSHYLSPAKF 222
Db      954 CARLQICSSCAQPRQPHDCCPQCSDCHEGERK-----YEPGESFQPG--ADPCEV 1002
Qy      223 -----SGPPKLKRNQAFSCISWVEVATCACLQAQGEAEQHCSPSLAEOM 270
Db      1003 CICEPQPEGPSPSLRCHRR--QCPS--LVGCPSPQLLPPG-----PQHCCPTCAEAL 1049

RESULT 9
US-10-189-971-6
; Sequence 6, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 6
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-6

Query Match      6.7%; Score 107.5; DB 15; Length 1057;
Best Local Similarity 20.0%; Pred. No. 0.11;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

Qy      2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMWTLVGR 53
Db      182 VSCPEKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPREPNCNLCTCLGG 237
Qy      54 LLP-----VPA-----KLNLIQVDNGALPSAVNGAAF----- 81
Db      238 FVTGRRRCPEPGCSHPLIPSGHCCPTCGCRYHGVTTASGETLPDPLDPTCSLCTCQEG 297
Qy      82 -----PSGPALQGGPKITLSGYCDFSFGDFSCNSCSNNLRHELFRKAICACLDNRN 134
Db      298 SMRCOKKPCAPALCPHPS--PGPCFC--PVCCHSCLSQGREHQ-----DG 337
Qy      135 EAFQPKMGK-----GRLGAALKR-----HSKGCNCR-SGCLK----- 166
Db      338 EEFEGPAGSCWCRCQAGVSCVRLQCPPLPKLQVTERGCCPRGCLAHGEHPEGSG 397
Qy      167 -----NYCEYE-----AKIMCSSICK-----CIACKNVEESPCKMLMSTP 203
Db      398 RWVPPDSACSCVCHGVVTCARLQICSSCAQPRQPHDCCPQCSDCHEGERK----- 450
Qy      204 HYMEPGD-FESSHYLSPAKF-----SGPPKLKRNQAFSCISWVEVATCACLQAQGE 255
Db      451 --YEPGESFQPG--ADPCEVICPEPQEPGPSPSLRCHRR--QCPS--LVGCPSPQLLPPG- 501
Qy      256 EAEQEHCSPSLAEOM 270
Db      502 ---PQHCCPTCAEAL 513

RESULT 10
US-10-189-971-16
; Sequence 16, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-16

Query Match      6.7%; Score 107.5; DB 15; Length 1251;
Best Local Similarity 20.0%; Pred. No. 0.14;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

Qy      2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMWTLVGR 53
Db      376 VSCPEKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPREPNCNLCTCLGG 431
Qy      54 LLP-----VPA-----KLNLIQVDNGALPSAVNGAAF----- 81
```

```
Db 432 FVTCGRRRCEPPGCSHPLIPSGCHCCTCGCRVHGVTASGETLPDPLDPTCSLCTCOEG 491
QY 82 -----PSGALOGPPKITLSGYCDQFSSGDFCNSCNCNLRHLERFKAIKACLDNRN 134
Db 492 SMRCOKKPCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 531
QY 135 EAFQPKMGK-----GRLGAALKR-----HSKGCNCKR-SGCLK-----166
Db 532 EEFEGPAGSCWCRCQAGQVSCVRLQCPPLCKLQVTERGSCCPCRCGLAHGEHPGS 591
QY 167 -----NYCEYE-----AKIMCSICK-----CIACKNYEESPERKMLMSTP 203
Db 592 RWPVPDSACSCVCHEGVVTCARIQCISSCAQPROGPHDCCPQCSDCEHEGRK-----644
QY 204 HYMEPGD-FESSHYLSPAKF-----SGPPKLRNRQAFSCISWEVVEATCACLLAQGE 255
Db 645 --YEPGESFQPG--ADPCEVCICEPQPEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 695
QY 256 EAEQHCSPSLAEQM 270
Db 696 ---POHCCPTCAEAL 707
```

RESULT 11

US-10-189-971-24

```
; Sequence 24, Application US/10189971
; Publication No. US20030028907A1
```

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Scoville, John

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleoti

TITLE OF INVENTION: Same

FILE REFERENCE: LEX-0360-USA

CURRENT FILING DATE: 2002-07-03

PRIOR FILING DATE: 2002-07-03

PRIOR FILING DATE: 2001-07-03

PRIOR FILING DATE: 2001-07-03

PRIOR FILING DATE: 2001-08-29

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 24

LENGTH: 1342

TYPE: PRT

ORGANISM: homo sapiens

US-10-189-971-24

```
Query Match
Best Local Similarity 20.0%; Pred. No. 0.15;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;
```

```
QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDOSSFPQSELKPKMTTLVGR 53
Db 467 VSCEPKACAPALCPFPARGDCPCDGCCE-----YLGESYLSNQEPDPREPNCNLTCLGG 522
QY 54 LLP-----VPA-----KLNLTITQVDNGALPSAVNGAAF-----81
Db 523 FVTCGRRRCEPPGCSHPLIPSGCHCCTCGCRVHGVTASGETLPDPLDPTCSLCTCOEG 582
QY 82 -----PSGALOGPPKITLSGYCDQFSSGDFCNSCNCNLRHLERFKAIKACLDNRN 134
Db 583 SMRCOKKPCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 622
QY 135 EAFQPKMGK-----GRLGAALKR-----HSKGCNCKR-SGCLK-----166
Db 623 EEFEGPAGSCWCRCQAGQVSCVRLQCPPLCKLQVTERGSCCPCRCGLAHGEHPGS 682
QY 167 -----NYCEYE-----AKIMCSICK-----CIACKNYEESPERKMLMSTP 203
Db 683 RWPVPDSACSCVCHEGVVTCARIQCISSCAQPROGPHDCCPQCSDCEHEGRK-----735
```

```
QY 204 HYMEPGD-FESSHYLSPAKF-----SGPPKLRNRQAFSCISWEVVEATCACLLAQGE 255
Db 736 --YEPGESFQPG--ADPCEVCICEPQPEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 786
QY 256 EAEQHCSPSLAEQM 270
Db 787 ---POHCCPTCAEAL 798
```

RESULT 12

US-10-189-971-10

Sequence 10, Application US/10189971

Publication No. US20030028907A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Scoville, John

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle

TITLE OF INVENTION: Same

FILE REFERENCE: LEX-0360-USA

CURRENT FILING DATE: 2002-07-03

PRIOR FILING DATE: 2002-07-03

PRIOR FILING DATE: 2001-07-03

PRIOR FILING DATE: 2001-08-29

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10

LENGTH: 1512

TYPE: PRT

ORGANISM: homo sapiens

US-10-189-971-10

```
Query Match
Best Local Similarity 20.0%; Pred. No. 0.17;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;
```

```
QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDOSSFPQSELKPKMTTLVGR 53
Db 637 VSCEPKACAPALCPFPARGDCPCDGCCE-----YLGESYLSNQEPDPREPNCNLTCLGG 692
QY 54 LLP-----VPA-----KLNLTITQVDNGALPSAVNGAAF-----81
Db 693 FVTCGRRRCEPPGCSHPLIPSGCHCCTCGCRVHGVTASGETLPDPLDPTCSLCTCOEG 752
QY 82 -----PSGALOGPPKITLSGYCDQFSSGDFCNSCNCNLRHLERFKAIKACLDNRN 134
Db 753 SMRCOKKPCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 792
QY 135 EAFQPKMGK-----GRLGAALKR-----HSKGCNCKR-SGCLK-----166
Db 793 EEFEGPAGSCWCRCQAGQVSCVRLQCPPLCKLQVTERGSCCPCRCGLAHGEHPGS 852
QY 167 -----NYCEYE-----AKIMCSICK-----CIACKNYEESPERKMLMSTP 203
Db 853 RWPVPDSACSCVCHEGVVTCARIQCISSCAQPROGPHDCCPQCSDCEHEGRK-----905
QY 204 HYMEPGD-FESSHYLSPAKF-----SGPPKLRNRQAFSCISWEVVEATCACLLAQGE 255
Db 906 --YEPGESFQPG--ADPCEVCICEPQPEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 956
QY 256 EAEQHCSPSLAEQM 270
Db 957 ---POHCCPTCAEAL 968
```

RESULT 13

US-10-189-971-12

Sequence 12, Application US/10189971

Publication No. US20030028907A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

```
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-2

Query Match      6.7%; Score 107.5; DB 15; Length 1628;
Best Local Similarity 20.0%; Pred No. 0.19;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY   2 VICOLKGGAML-----CIDNCGARELKALHLLPYDDQSSFFQSSELPKPMTTLVGR 53
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   753 VSCPRKACAPALCPFPARGDCPCDGC-----VLGESYLSNQEFPPPREPCNLCTCLG 808

QY   54 LLP-----VPA-----KUNLITQVDNGALPSAVNCAAF-----81
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   809 FVTGRRPCEPPGCSHPLIPSGHCCTCOGCRTHGVTTASGETLPDLPTCSILCTCQG 868

QY   82 -----PSGPALQQGPKITLSGYCDCFSSGDFCNSCCNNLRHELRFKAIRACLDRNP 134
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   869 SMCQKKPCAPALCPHP-----PGPCFC-----PVCHSCLSQGREHQ-----DG 908

QY   135 EAFQPKMGK-----GRIGAALKR-----HSKGCKNKR--SGCLK-----166
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   909 EEFEAGPAGSCWRCQAGQVSCVRLQCPPLPCKLQVTERGCCPCRGCIAHGEEHPGS 968

QY   167 -----NYCBYE-----AKIMSSICK-----CIACKNVESPERKMMLMSTP 203
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   969 RWPVPDSACSVCHEGVVTCARTIQITSSCAQPRQPHDCQPQCSDCEHEGRK-----1021

QY   204 HYMEPGD-FESSHYLSPAKE-----SGPPKLRKNRAFPSCISWEVVTEATCACLQAQE 255
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   1022 --YEPGESFQPG--ADPCEVCICEPOEGPPSLRCHRR--OCPS--LVGCPPSOLLPPG- 1072

QY   256 EAEQEHCSPLAQDM 270
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   1073 ---PQHCCPTCAEAL 1084


RESULT 15
US-10-189-971-22
; Sequence 22, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; TITLE OF INVENTION: Turner, C. Alexander Jr.
; FILE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 759
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-22

Query Match      6.3%; Score 100; DB 15; Length 759;
Best Local Similarity 22.6%; Pred. No. 0.38;
Matches 56; Conservative 24; Mismatches 70; Indels 98; Gaps 16;

QY   82 PSGPALQQGPKITLSGYCDCFSSGDFCNSCCNNLRHELRFKAIRACLDRNPFAQPKM 141
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   7 PCAPALCPHP-----PGPCFC-----PVCHSCLSQGREHQ-----DGEFEFGPA 46

QY   142 GK-----GRIGAALKR-----HSKGCKNKR--SGCLK-----166
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   47 GSCEWCRCQAGVSCVRLQCPPLPCKLQVTERGCCPCRGCIAHGEEHPGSRWVPDS 106
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:34:58 ; Search time 20.362 Seconds
(without alignments)
612.991 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLHIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued_Patents_AA.*
- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pap.*
 - 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pap.*
 - 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pap.*
 - 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pap.*
 - 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pap.*
 - 6: /cgn2_6/ptodata/1/1aa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	93.5	5.8	652	2	US-08-751-305-2	Sequence 2, Appli
2	92.5	5.8	921	4	US-09-699-286A-9	Sequence 9, Appli
3	87.5	5.5	610	1	US-08-365-470-3	Sequence 3, Appli
4	87.5	5.5	610	3	US-09-209-668-19	Sequence 19, Appl
5	87.5	5.5	610	4	US-09-009-490A-89	Sequence 89, Appl
6	87.5	5.4	902	4	US-09-699-266A-12	Sequence 12, Appl
7	86.5	5.4	418	3	US-08-795-430-13	Sequence 13, Appl
8	86.5	5.4	418	4	US-09-355-700-13	Sequence 13, Appl
9	86.5	5.4	610	6	5217870-2	Patent No. 5217870
10	86.5	5.4	830	1	US-08-110-158-4	Sequence 4, Appli
11	85.5	5.3	769	2	US-08-789-078-1	Sequence 1, Appli
12	85.5	5.3	769	2	US-08-752-633-1	Sequence 1, Appli
13	85.5	5.3	769	2	US-08-476-062A-45	Sequence 45, Appl
14	85.5	5.3	769	2	US-07-728-215-31	Sequence 31, Appl
15	85.5	5.3	769	4	US-08-938-085A-31	Sequence 31, Appl
16	85.5	5.3	769	4	US-10-072-844-31	Sequence 31, Appl
17	85.5	5.3	769	5	PCT-US95-04886-1	Sequence 1, Appli
18	85.5	5.3	769	5	PCT-US96-01314-45	Sequence 45, Appl
19	85.5	5.3	217	4	US-09-602-543-5	Sequence 5, Appli
20	85.5	5.3	235	4	US-09-602-543-4	Sequence 4, Appli
21	85.5	5.3	676	3	US-08-630-172-10	Sequence 10, Appl
22	85.5	5.3	676	3	US-09-375-419-10	Sequence 10, Appl
23	84.5	5.3	696	3	US-08-899-437-23	Sequence 23, Appl
24	84.5	5.3	696	3	US-09-126-121-23	Sequence 23, Appl
25	84.5	5.3	720	3	US-08-899-437-6	Sequence 6, Appli
26	84.5	5.3	720	3	US-09-126-121-6	Sequence 6, Appli
27	83.5	5.2	228	4	US-09-431-888-7	Sequence 7, Appli

28	83.5	5.2	350	2	US-08-999-811-4	Sequence 4, Appli
29	83.5	5.2	350	2	US-08-824-996-2	Sequence 2, Appli
30	83.5	5.2	350	3	US-09-042-105-4	Sequence 4, Appli
31	83.5	5.2	350	3	US-08-510-133A-33	Sequence 33, Appl
32	83.5	5.2	350	3	US-08-585-895-33	Sequence 33, Appl
33	83.5	5.2	419	2	US-08-999-811-2	Sequence 2, Appli
34	83.5	5.2	419	3	US-09-042-105-2	Sequence 2, Appli
35	83.5	5.2	419	3	US-09-042-105-18	Sequence 18, Appl
36	83.5	5.2	419	3	US-08-795-430-8	Sequence 8, Appli
37	83.5	5.2	419	3	US-08-510-133A-35	Sequence 35, Appli
38	83.5	5.2	419	4	US-09-355-700-8	Sequence 8, Appli
39	83.5	5.2	419	4	US-09-355-700-58	Sequence 58, Appl
40	83.5	5.2	419	4	US-08-601-132-33	Sequence 33, Appl
41	83.5	5.2	419	4	US-08-706-054A-3	Sequence 3, Appli
42	83.5	5.2	419	4	US-09-313-299-3	Sequence 3, Appli
43	83.5	5.2	419	5	PCT-US96-09001-2	Sequence 2, Appli
44	81.5	5.1	1964	4	US-09-467-997-1	Sequence 1, Appli
45	81	5.1	415	3	US-08-795-430-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-751-305-2
; Sequence 2, Application US/08751305
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Fenner et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-305-2

Query Match	5.8%;	Score 93.5;	DB 2;	Length 652;
Best Local Similarity	27.0%;	Pred No. 0.84;	Mismatches 51;	Indels 19;
Matches	30;	Conservative 11;	Length 652;	
QY	96	SGYC--DCFSGGDFCNSCNLNRKELERFKAKACLDNRNPEAFQPKMGKRLGAALRH	153	
Db	268	NGGCHQDCFEFGDGLGCG---RPGFRLDLLVTCA SRNPCSSSPCRG-----	313	
QY	154	SKGCKNRSGGLKNY-CECYEAKIMCSCICKIACKNYEESPERKMLMSTP	203	
Db	314	--GATCVLGPHGKNTCRCPQGYQLDSSQLDCVDDECQDSPCAQECVNTP	362	

	US-08-795-430-13	
	; Sequence 13, Application US/08795430	
	; Patent No. 6130071	
	GENERAL INFORMATION:	
	APPLICANT: Alitalo, Kari	
	APPLICANT: Joukov, Vladimir	
	TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)	
	TITLE OF INVENTION: Protein and Gene, Mutants Thereof, and Uses Thereof	
	NUMBER OF SEQUENCES: 57	
	CORRESPONDENCE ADDRESS:	
	ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun	
	STREET: 6300 Sears Tower, 233 South Wacker Drive	
	CITY: Chicago	
	STATE: Illinois	
	COUNTRY: United States of America	
	ZIP: 60606-6402	
	COMPUTER READABLE FORM:	
	MEDIUM TYPE: Floppy disk	
	COMPUTER: IBM PC compatible	
	OPERATING SYSTEM: PC-DOS/MS-DOS	
	SOFTWARE: PatentIn Release #1.0, Version #1.30	
	CURRENT APPLICATION DATA:	
	APPLICATION NUMBER: US/08/795,430	
	FILING DATE: 01-AUG-1995	
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER: 08/340,011	
	FILING DATE: 14-NOV-1994	
	ATTORNEY/AGENT INFORMATION:	
	NAME: Gass, David A.	
	REGISTRATION NUMBER: 38,153	
	REFERENCE/DOCKET NUMBER: 28967/33691	
	TELEPHONE: 312/474-6300	
	TELEFAX: 312/474-0448	
	TELEX: 25-3856	
	INFORMATION FOR SEQ ID NO: 13:	
	SEQUENCE CHARACTERISTICS:	
	LENGTH: 418 amino acids	
	TYPE: amino acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: protein	
	US-08-795-430-13	
	Query Match	5.4%; Score 86; DB 3; Length 418;
	Best Local Similarity	18.6%; Pred. No. 2.6;
	Matches	48; Conservative 24; Mismatches 90; Indels 96; Gaps 10;
QY	14 CIDNCGAREKALHLPYDDGSSP-----PQSELPKPMW-----	48
DB	161 CGGCCNSEGLQMISTNTYSIKTLFTIVPLSHGPKPTVSFANHTSCRCMSKLDVYRQV	220
QY	49 -TLVGRLLPVPAK-----LNLITVDNGALPSAVNGAAPPSPALQ	88
DB	221 HSIIRSLPATQTQCHVANKTCPKNHVNQCICLAHQDFG-FSSHIGDSDTSEGFHC	279
QY	89 GPPKITLSGYCDFFSGDF-----CN-----SCSC-----NNL-----	116

RESULT 9

; LENGTH: 610

RESULT 10

ADDRESSEE: Patrea L. Pabst

Query Match	5.48;	Score 86;	DB 1;	Length 830;
-------------	-------	-----------	-------	-------------

RESULT 11

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,078
FILING DATE: 03-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,513
FILING DATE: 19-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label= cyto
OTHER INFORMATION: /note= "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
AUTHORS: Power,
TITLE: LFA-1 Amino acid sequence (B2) (from human
TITL: tonsil)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769

```
US-08-789-078-1

Query Match          5.3%; Score 85.5; DB 2; Length 769;
Best Local Similarity 23.4%; Pred. No. 6.4;
Matches      32; Conservative    13; Mismatches   37; Indels     55; Gaps       10;
```

QY 90 PPKITISGVDCDFS-----SGD-----FCNSCSN----NLRLHELERFKAKAC 129
| | : ||| : | : | | | | : | : |
Db 525 PGKLIVGYCECDPTINCERYNGVCGPGRGLCFGCKRCRCHPFGFSACOCER--TTTEGC 582

QY 130 LDRNPFAFPKMKGRIALGAAKLHRSKCNCNRSGCLKNYCCEYEAKIM-----CSSI 181
| | : | | | : | | | | | | : | | |
Db 583 L--NPRRVES-GRGR-----CRC-----NVCECHSGYLPLQCPCGPCSP 621

QY 182 C-----KCIACKNYEESP 194
| | : | | | : | | | | | : | | |
Db 622 CGKIYSIAECKLKEGP 638

RESULT 12
US-08-752-633-1
Sequence I, Application US/08752633
Patent No. 5863889
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,633
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496


```

; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 541..581
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 582..617
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 701..723
; OTHER INFORMATION: /label= trans
; OTHER INFORMATION: /note= "transmembrane domain"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 724..769
; OTHER INFORMATION: /label= cyto
; OTHER INFORMATION: /note= "cytoplasmic domain"
; PUBLICATION INFORMATION:
; AUTHORS: Pigott,
; TITLES: LFA-1 Amino acid sequence (B2) (from human
; TITLE: tonsil)
; JOURNAL: The Adhesion Molecule Facts Book
; PAGES: 96-96
; DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
US-08-752-633-1

```

Query Match	5.3%	Score 85.5;	DB 2;	Length 769;
Best Local Similarity	23.4%;	Pred. No. 6.4;		
Matches	32;	Conservative 13;	Mismatches 37;	Indels 55; Gaps 10;
Qy	90	PKKITLSGGYCDPFS-----SGD-----FCNSCSGN-----NLRHLEFRFAIKAC	129	
Db	525	PGKLIIGQYCECDTINCERYNGOVGGPGRLGFCGKRCRHPFGSACOCER--TTEGC	582	
Qy	130	LDRNPEAFQPKMGKGLGAALKRHSKGCNCKRSGLCNKYCEYEAKIM-----CSSI	181	
Db	583	L--NPRVCS-GRGR-----CRC-----NVCECHSGVQLPLCQECPCGCPSP	621	
Qy	182	C-----KCIACKNYDESP	194	
Db	622	CGKYISCAECLKFKPKGP	638	

```

RESULT 13
US-08-476-062A-45
; Sequence 45, Application US/08476062A
; Patent No. 5877275
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
; NUMBER OF SEQUENCES: 53
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston.
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,062A
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,081
FILING DATE: 21-MAR-1994
APPLICATION NUMBER: 07/637,830
FILING DATE: 04-JAN-1991
APPLICATION NUMBER: 07/539,842
FILING DATE: 18-JUN-1990
APPLICATION NUMBER: 07/212,573
FILING DATE: 28-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/068003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-476-062A-45

```

Query Match	5.3%	Score 85.5;	DB 2;	Length 769;
Best Local Similarity	23.4%;	Pred. No. 6.4;		
Matches	32;	Conservative 13;	Mismatches 37;	Indels 55; Gaps 10;
QY	90	PKKITLSGYGDCDFS-----SGD-----FCNSCSGN-----NLRHELERFKATKAC	129	
Db	525	PGRLIYQYCECDTINCERYNGVCGGPGRLGFCGKRCRPHFGSACQCR-----TTEGC	582	
QY	130	LDRNPEAFQPKMGKGRGLGAALRHSKGCNKRSGCLKNKYCECYEAKTM-----CSSI	181	
Db	583	L--NPRVECS-GRGR-----CRC-----NVCECHSGYOLPLCQECPCGCPSP	621	
QY	182	C-----KCIACKNYEESP	194	
Db	622	CGKYISCAECLKFKEKP	638	

```

RESULT 14
US-07-728-215-31
; Sequence 31, Application US/07728215
; Patent No. 5962643
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytel, Robert
; TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:22:12 ; Search time 59.0993 seconds
(without alignments)
792.299 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGGAQLCIDNCGA.....GRCLSQILHIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_13Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	100.0	295	21	AA1980
2	1351	78.2	299	21	AA1981
3	639	40.0	528	24	ABG74678
4	435	27.2	950	22	ABG62035
5	396	24.8	438	20	AAW83392
6	362	22.6	280	22	ABG17958
7	322	20.1	147	22	ABG6025
8	322	20.1	147	22	AAW93330
9	248	15.5	53	22	AAW76158

10	223.5	14.0	609	24	ABP70527	Histone deacetylase
11	186	11.6	243	22	ABG68888	Drosophila melanog
12	165	10.3	251	22	AAW93348	Human polyprotein
13	115	7.2	1192	24	ABP97376	Human kiellin-like
14	115	7.2	1207	24	ABP97377	Human kiellin-like
15	115	7.2	1477	24	ABP97371	Human kiellin-like
16	115	7.2	1535	24	ABP97374	Human kiellin-like
17	115	7.2	1593	24	ABP97369	Human kiellin-like
18	115	7.2	1774	23	ABG69800	Human REWAP-6 prot
19	110.5	6.9	732	18	AAW25642	Human RECK cancer
20	107.5	6.7	1057	24	ABP97370	Human kiellin-like
21	107.5	6.7	1251	24	ABP97375	Human kiellin-like
22	107.5	6.7	1342	24	ABP97379	Human kiellin-like
23	107.5	6.7	1512	24	ABP97372	Human kiellin-like
24	107.5	6.7	1570	24	ABP97373	Human kiellin-like
25	107.5	6.7	1628	24	ABP97368	Human kiellin-like
26	100	6.3	759	24	ABP97378	Human kiellin-like
27	99	6.2	2931	22	ABG68229	Drosophila melanog
28	97	6.1	602	21	AAW95660	Human ZntR2 protei
29	97	6.1	1987	22	ABG61083	Drosophila melanog
30	96.5	6.0	359	21	AAW09497	Arabidopsis thalia
31	96.5	6.0	443	21	AAW09496	Arabidopsis thalia
32	96.5	6.0	472	21	AAW09495	Arabidopsis thalia
33	96.5	6.0	1788	22	ABG62995	Drosophila melanog
34	96.5	6.0	4547	22	ABG59051	Drosophila melanog
35	96	6.0	773	22	ABG59751	Drosophila melanog
36	96	6.0	3635	23	ABG81589	Mouse laminin alph
37	96	6.0	3635	23	AAW50357	Mouse laminin-15 a
38	95.5	6.0	1035	22	ABG66062	Drosophila melanog
39	94.5	5.9	538	24	ABU12120	Human protein modi
40	94.5	5.9	577	24	ABU12118	Human LP protein L
41	94.5	5.9	759	23	AAU97037	Human laminin 8 po
42	94	5.9	1792	21	AAW48443	Human laminin 8 po
43	94	5.9	1800	21	AAW48445	Human laminin 8 po
44	94	5.9	1816	21	AAW48442	Human laminin 8 po
45	94	5.9	1824	21	AAW48444	Human laminin 8 po

ALIGNMENTS

RESULT 1
AA1980
ID AAY68463 standard; Protein; 295 AA.

XX AAY68463;

AC AAY68463;

DT 25-APR-2000 (first entry)

XX Mouse testis specific factor tesmin SEQ ID NO:4.

XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.

XX Mus musculus.

XX WO200004147-A1.

PD 27-JAN-2000.

PF 16-JUL-1999; 99WO-JP03859.

PR 17-JUL-1998; 98JP-0219856.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

DR WPI: 2000-147785/13.

DR N-PSDB; AA288155, AA288157.

XX

PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 useful for investigation of germ cell differentiation and sterility -
 Claim 1; Page 47-49; 63pp; Japanese.
 The present sequence represents a male germ cell regulatory factor
 expressed specifically in spermatocytes, designated tesmin. Tesmin
 can be used in the investigation of the mechanisms of germ cell
 differentiation and sterility.
 SQ Sequence 295 AA;

Query Match 100.0%; Score 1599; DB 21; Length 295;
 Best Local Similarity 100.0%; Pred. No. 3e-144;
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVICQLKGGAGMLCINCGARELKALHLLPQYDDQSSFPQSELPKPMTTLVGRLLPVPK 60
 Db 1 MVICQLKGGAGMLCINCGARELKALHLLPQYDDQSSFPQSELPKPMTTLVGRLLPVPK 60
 Qy 61 LNLITQVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGFCFSSGDFCNSCENLRLHEL 120
 Db 61 LNLITQVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGFCFSSGDFCNSCENLRLHEL 120
 Qy 121 ERFAIKACLDNRNPEAFQPKMGKGRGALHRSKGCNCRKRGCLKNYCEYEAKIMCSS 180
 Db 121 ERFAIKACLDNRNPEAFQPKMGKGRGALHRSKGCNCRKRGCLKNYCEYEAKIMCSS 180
 Qy 181 IKCKIAKNYEESEPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRNQRQAFSCISW 240
 Db 181 IKCKIAKNYEESEPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRNQRQAFSCISW 240
 Qy 241 EWEATCACLIAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 Db 241 EWEATCACLIAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295

RESULT 2
 AAY68464
 ID AAY68464 standard; Protein; 299 AA.
 AC AAY68464;
 XX
 XX
 XX 25-APR-2000 (first entry)
 DT
 DE Human testis specific factor tesmin SEQ ID NO:5.
 XX
 XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility.
 XX
 XX Homo sapiens.
 OS
 XX WO200004147-A1.
 PN
 XX 27-JAN-2000.
 PD
 XX 16-JUL-1999; 99WO-JP03859.
 PF
 XX 17-JUL-1998; 98JP-0219856.
 PR
 XX (CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX (AGENCY) AGENCY OF IND SCI & TECHNOLOGY.
 PA
 XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 PI
 XX WPI; 2000-147785/13.
 DR N-PSDB; AAZ88156.
 DR
 PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 useful for investigation of germ cell differentiation and sterility -
 Claim 1; Page 50-52; 63pp; Japanese.

XX The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX SQ Sequence 299 AA;

Query Match 78.2%; Score 1251; DB 21; Length 299;
 Best Local Similarity 76.3%; Pred. No. 5.3e-111;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICQLKGGAGMLCINCGARELKALHLLPQYDDQSSFPQSELPKPMTTLVGRLLPVPK 60
 Db 1 MVICQLKGGAGMLCINCGARELKALHLLPQYDDQSSFPQSELPKPMTTLVGRLLPVPK 60
 Qy 61 LNLITQVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGFCFSSGDFCNSCENLRLHEL 116
 Db 61 LNLITQVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGFCFSSGDFCNSCENLRLHEL 120
 Qy 117 RHELEKAIKACLDNRNPEAFQPKMGKGRGALHRSKGCNCRKRGCLKNYCEYEAKI 176
 Db 121 HHDIERFAIKACLDNRNPEAFQPKMGKGRGALHRSKGCNCRKRGCLKNYCEYEAKI 180
 Qy 177 MCSICKIAKNYEESEPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRNQRQAFS 236
 Db 181 MCSICKIAKNYEESEPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRNQRQAFS 240
 Qy 237 CISWEVVEATCACLIAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 Db 241 CISWEVVEATCACLIAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKME 299

RESULT 3
 ABG74678
 ID ABG74678 standard; Protein; 528 AA.
 XX
 XX AC ABG74678;
 XX
 XX 10-MAY-2003 (first entry)
 DT
 DE Human CGDD protein 2369279CD1 SEQ ID 4.
 XX
 XX Human; cell growth, differentiation and death; CGDD; cardiact; cancer;
 KW cytosatic; neuroprotective; gene therapy; diagnosis; cardiovascular;
 KW neurological disorder.
 XX
 XX Homo sapiens.
 OS
 XX WO2003014322-A2.
 PN
 XX 20-FEB-2003.
 PD
 XX 08-AUG-2002; 2002WO-US25465.
 PF
 XX 08-AUG-2001; 2001US-311017P.
 PR 17-AUG-2001; 2001US-313070P.
 PR 17-AUG-2001; 2001US-313071P.
 PR 24-AUG-2001; 2001US-314678P.
 PR 31-AUG-2001; 2001US-316692P.
 PR 07-SEP-2001; 2001US-317913P.
 PR 14-SEP-2001; 2001US-322182P.
 PR 07-DEC-2001; 2001US-340747P.
 PR 20-DEC-2001; 2001US-342761P.
 PR 29-MAR-2002; 2002US-369129P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Azimzai Y, Barroso I, Baughn MR, Becha SD, Borowsky ML, Duggan BM;
 PI Elliott VS, Emerling BM, Forsythe IJ, Gietzen KJ, Gorvad AB;
 PI Graul RC, Griffin JA, Gururajan R, Hafalia AJA, Ison CH, Kable AB;
 PI Khan FA, Lee S, Lee SY, Li JX, Reddy R, Richardson TW;
 PI Sprague WW, Swarnakar A, Tang YT, Warren BA, Xu Y, Yao MG, Yue H;

XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic.
OS Homo sapiens.
XX WO200153317-A2.
PN 02-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01329.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.

```

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483232/52.
DR
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer
XX
XX Claim 11; SEQ ID NO 1409; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a protein of the
CC invention.
XX
XX SQ Sequence 147 AA;
    Query Match          20.1%; Score 322; DB 22; Length 147;
    Best Local Similarity 73.8%; Pred. No. 1.le-22;
    Matches 62; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
Oy 1 MVICOLKGGAMLCIDNCGARELKHLLPOYDQSFPOSELPKPMTTLVGRLLPVPK 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 MVICQLKGGTQMLCIDNSTRRELKALHLVPOYQDQNNYLOSDVPKPMTALVGRFLPASTK 119
Oy 61 LNLITVDNGALPSAVNGAAPPSSG 84
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 LNLITQOLEGALPSVNGSAPPSSG 143

RESULT 8
AAM95330
ID AAM95330 standard; Protein; 147 AA.
XX
AC AAM95330;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen SEQ ID NO: 3988.
DE
XX Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
XX
XX Homo sapiens.
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190075.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 07-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 11-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.

```



```
XX 10-MAY-2003 (first entry)
XX Human kielin-like protein NHP10, SEQ ID NO:20.
XX Human; kielin-like protein; NHP; chromosome 7; development;
XX signal transduction; kielin-related disorder; diagnosis; treatment;
XX drug screening; clinical trial monitoring; cosmetic; nutraceutical.
XX Homo sapiens.
XX WO2003004609-A2.
XX 16-JAN-2003.
XX 03-JUL-2002; 2002WO-US21118.
XX 03-JUL-2001; 2001US-302949P.
XX 29-AUG-2001; 2001US-315634P.
XX (LEXI-) LEXICON GENETICS INC.
XX Walke DW, Scoville J, Turner CA;
XX WPI; 2003-210352/20.
XX N-PSDB; ABZ75048.
XX New human kielin-like nucleic acids, useful for preparing a composition
XX for treating human kielin-related disorders -
XX Claim 3; Page 75-77; 85pp; English.
XX The invention relates to novel human kielin-like proteins (referred to as
XX NHPs; ABP97368-ABP97379) and cDNAs encoding them (ABZ75039-ABZ75050). The
XX invention also encompasses recombinant expression vectors and host cells
XX comprising a nucleic acid of the invention. The NHP cDNAs were obtained
XX by aligning cDNAs from human kidney, foetal kidney, prostate and lymph
XX node mRNAs. The NHPs of the invention are all apparently encoded on
XX chromosome 7 and share sequence similarity with animal kielin proteins.
XX Kielines are secreted proteins implicated in a number of biological
XX processes such as development and signal transduction. The NHP nucleic
XX acids and proteins may be used in the diagnosis and treatment of human
XX kielin-related disorders, and may also be used in drug screening and
XX clinical trial monitoring. They may further be used in cosmetic and
XX nutraceutical applications. The present sequence represents an NHP
XX of the invention.
XX Sequence 1207 AA;
XX
XX Query Match 7.2%; Score 115; DB 24; Length 1207;
XX Best Local Similarity 21.1%; Pred. No. 0.097;
XX Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;
XX
XX 2 VICOLKGAQWL-----CIDNCGARELKALHLLPQYDDQSSFPQSELPKPMVTLVGR 53
XX 367 VSCPEKACALCPFPARGDCPCDCE-----YLGESYLSNQEFPPREPNCNLTCLGG 422
XX
XX 54 LLPVPAKLNLITQVDNGALPFAVNGAAPPSPALQGPPIKTLGYC-----DC-----F 102
XX 423 F-----VTCGRRRCEPPGCSHPLIP-----SGHCCTCGGCRYHGVTT 460
XX
XX 103 SSGD-----FCNSCSCNNLRHELFRKAKACLDNRNPEAFQPKMGK-----GRL 146
XX 461 ASGETLPDLPDPTCSLCTCGREHQ-----DGEFEGPAGSCWCRCQAGQV 507
XX
XX 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCEYE----- 173
XX 508 SCVRLOCPPLPKLQVTRGSCRCRCGLAHGEHPEGRWVPPDSACSCVCHEGVTT 567
XX 174 -AKIMCSICK-----CIACKNYEESPERKMLMSTPHYMEPPGD-FESSHYLSPAKE 222
XX 568 CARIQCISSCAQPRQPHDCCPQCSQCEHEGRK-----YEPGESFQPG-ADCEV 616
XX
```

```
QY 223 -----SGPPKLRKNRQAFSCISWEVVEATCATCLLAQGEAEQEHCSPSLAEQM 270
DB 617 CICEPQEPGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----POHCPTCAEAL 663

RESULT 15
ABP97371
ID ABP97371 standard; Protein; 1477 AA.
XX AC ABP97371;
XX 10-MAY-2003 (first entry)
XX Human kielin-like protein NHP4, SEQ ID NO:8.
XX Human; kielin-like protein; NHP; chromosome 7; development;
XX signal transduction; kielin-related disorder; diagnosis; treatment;
XX drug screening; clinical trial monitoring; cosmetic; nutraceutical.
XX Homo sapiens.
XX WO2003004609-A2.
XX 16-JAN-2003.
XX 03-JUL-2002; 2002WO-US21118.
XX 03-JUL-2001; 2001US-302949P.
XX 29-AUG-2001; 2001US-315634P.
XX (LEXI-) LEXICON GENETICS INC.
XX Walke DW, Scoville J, Turner CA;
XX WPI; 2003-210352/20.
XX N-PSDB; ABZ75042.
XX New human kielin-like nucleic acids, useful for preparing a composition
XX for treating human kielin-related disorders -
XX Claim 3; Page 50-53; 85pp; English.
XX The invention relates to novel human kielin-like proteins (referred to as
XX NHPs; ABP97368-ABP97379) and cDNAs encoding them (ABZ75039-ABZ75050). The
XX invention also encompasses recombinant expression vectors and host cells
XX comprising a nucleic acid of the invention. The NHP cDNAs were obtained
XX by aligning cDNAs from human kidney, foetal kidney, prostate and lymph
XX node mRNAs. The NHPs of the invention are all apparently encoded on
XX chromosome 7 and share sequence similarity with animal kielin proteins.
XX Kielines are secreted proteins implicated in a number of biological
XX processes such as development and signal transduction. The NHP nucleic
XX acids and proteins may be used in the diagnosis and treatment of human
XX kielin-related disorders, and may also be used in drug screening and
XX clinical trial monitoring. They may further be used in cosmetic and
XX nutraceutical applications. The present sequence represents an NHP
XX of the invention.
XX Sequence 1477 AA;
XX
XX Query Match 7.2%; Score 115; DB 24; Length 1477;
XX Best Local Similarity 21.1%; Pred. No. 0.13;
XX Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;
XX
XX 2 VICOLKGAQWL-----CIDNCGARELKALHLLPQYDDQSSFPQSELPKPMVTLVGR 53
XX 637 VSCPEKACALCPFPARGDCPCDCE-----YLGESYLSNQEFPPREPNCNLTCLGG 692
XX
XX 54 LLPVPAKLNLITQVDNGALPFAVNGAAPPSPALQGPPIKTLGYC-----DC-----F 102
XX 693 F-----VTCGRRRCEPPGCSHPLIP-----SGHCCTCGGCRYHGVTT 730
XX 103 SSGD-----FCNSCSCNNLRHELFRKAKACLDNRNPEAFQPKMGK-----GRL 146
XX
```

```
Db 731 ASGETLPDPLDPTCSLCTCOGREHQ-----DGEFEFGPAGSCWCRCAQGV 777
QY 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCEVE----- 173
Db 778 SCVRLQCPPLPCKLQVTERGSCCPRCCLAHGEEHPEGSRWVPPDSACSSCVCHGVWT 837
QY 174 -AKINCSSICK-----CIACKNYEESPERKMLMSTPHYMEPGD-FESSHYLSPAKF 222
Db 838 CARIQCISSCAQPRQGHDCPCQCSDCHEGRK-----YEPGESFQPG--ADPCEV 886
QY 223 -----SGPPKLRKNRQAFSCISWEVVVEATCACLQAQGEAEQEHCSFSLAEQM 270
Db 887 CICEPQPEGPESLRCHRR--QCPS--LVGCPPSOLLPPG-----PQHCCPTCAEAL 933
```

Search completed: July 24, 2003, 13:36:57
Job time : 51.0993 secs


```
Db 181 MVICQLKGGQMLCIDNCGARELKALHLLPYDDQSSFFQSELPKPMPTTLVGRLLPVPAK 240
Qy 61 LNLITQOLEGALPSVNGSAPPSGSLTPGPKITLAGYCDGCFASGDFCNCCNCCNNL 120
Db 241 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLISGYCDGCFSSGDFCNCS-----CNL 296
Qy 121 HHDIERFAIKACLDNRPEAFQPKIGKQGLGNVPQHNGKNCRRSGCLNKCCEYBAQI 180
Db 297 RHELEERFAIKACLDNRPEAFQPKMGKGRUGAALRHSGKNCCKRSGLNKCCEYBAKI 356
Qy 181 MCSICICGCKNYEESPERKTLMSPNYMTQGLESHYLPPTKFGSLRFRSHDRPSS 240
Db 357 MCSICICGCKNYEESPERKTLMSPNYMTQGLESHYLPPTKFGSLRFRSHDRPSS 240
Qy 241 CISWEVEATCACLAGEAEAEKHCSCLAEOIMLEEFGRCLSQILHTEFKSKGLKME 299
Db 417 CISWEVEATCACLAGEAEAEKHCSPSLAEQIMLEEFGRCLSQILHTEFKSKGLKIE 475

RESULT 2
OBVIE1
ID 08VIE1 PRELIMINARY; PRT; 475 AA.
AC 08VIE1
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE Tesmin.
GN TSM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Chinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB057422; BAB64934.1; -.
DR InterPro: IPR005172; CAC.
DR Pfam: PF03638; CXC; 1.
SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Query Match 75.5%; Score 1244; DB 11; Length 475;
Best Local Similarity 75.9%; Pred. No. 3.9e-120;
Matches 227; Conservative 25; Mismatches 43; Indels 4; Gaps 1;

Qy 1 MVICQLKGGQMLCIDNCGARELKALHLLPYDDQSSFFQSELPKPMPTTLVGRLLPVPAK 240
Db 181 MVICQLKGGQMLCIDNCGARELKALHLLPYDDQSSFFQSELPKPMPTTLVGRLLPVPAK 240
Qy 61 LNLITQOLEGALPSVNGSAPPSGSLTPGPKITLAGYCDGCFASGDFCNCCNCCNNL 120
Db 241 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLISGYCDGCFSSGDFCNCS-----CNL 296
Qy 121 HHDIERFAIKACLDNRPEAFQPKIGKQGLGNVPQHNGKNCRRSGCLNKCCEYBAQI 180
Db 297 RHELEERFAIKACLDNRPEAFQPKMGKGRUGAALRHSGKNCCKRSGLNKCCEYBAKI 356
Qy 181 MCSICICGCKNYEESPERKTLMSPNYMTQGLESHYLPPTKFGSLRFRSHDRPSS 240
Db 357 MCSICICGCKNYEESPERKTLMSPNYMTQGLESHYLPPTKFGSLRFRSHDRPSS 240
Qy 241 CISWEVEATCACLAGEAEAEKHCSCLAEOIMLEEFGRCLSQILHTEFKSKGLKME 299
Db 417 CISWEVEATCACLAGEAEAEKHCSPSLAEQIMLEEFGRCLSQILHTEFKSKGLKIE 475

RESULT 3
Q8C107
ID 08C107 PRELIMINARY; PRT; 419 AA.
```

```
AC 08C107;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Weakly similar to TESMIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL: AK029288; BAC26373.1; -.
SQ SEQUENCE 419 AA; 45301 MW; 24E3D4D454217D5B CRC64;

Query Match 40.6%; Score 669.5; DB 11; Length 419;
Best Local Similarity 45.5%; Pred. No. 1e-60;
Matches 150; Conservative 38; Mismatches 85; Indels 57; Gaps 10;

Qy 10 TMLCIDNSRTRELKALHLPYQYQDONNYLQSD---VPKPMALVG-RFLPASTKLNLTIT 65
Db 94 SQIVTTSQPOORLIMPATPLPQIPNLNPPGTVLAPACTGNVGYAVLPA---QYVT 149
Qy 66 QQLLEGALPSVNGSAPF-----PSGST-LPGPP-----KITLAGYCDC 101
Db 150 LQOQSSVYSIASNSNFTGTSGIQTQARLPFNGLIIPSESTRPKPCNCTKSLCLKYCDC 209
Qy 102 PASGDFCNCCNCCNHLHDIERFAIKACLDNRPEAFQPKIGKQGLGNVPQHNGK 161
Db 210 FANGFCNCCNCCNCCNHLHDIERFAIKACLDNRPEAFQPKIGKQGLGNVPQHNGK 269
Qy 162 NCRSGCLNKCCEYBAQIMCSCICKGCKNYEESPERKTLMSPN-----YMQTGG 215
Db 270 NCRSGCLNKCCEYBAQIMCSCICKGCKNYEESPERKTLMSPN-----YMQTGG 328
Qy 216 EGSYLPPTKFSG-----LPR-----FSDHRRSSCISWEVEATCACLAGEAEKE 264
Db 329 -----KTLSSQISDLTRPTPALMSAGKLPFTFTEVAEATCNLLAQAEQADKK 381
Qy 265 HCSKCLAEQIMLEEFGRCLSQILHTEFKSK 294
Db 382 GKSKAAERIMLEEFGRCLMSVINSACKAK 411

RESULT 4
Q9V6Q8
ID Q9V6Q8 PRELIMINARY; PRT; 950 AA.
AC Q9V6Q8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE CG6061 protein.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
```



```

RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015732; BAB29949.1; -.
SQ SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;

Query Match 22.4%; Score 369; DB 11; Length 277;
Best Local Similarity 73.2%; Pred. No. 8.1e-30;
Matches 71; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 1 MVICOLGGTOMCIDNSRTRELKALHLVPOYQONNYLQSDVPKPMYALVGRFLPASTK 60
DB 181 MVICOLGGAQMLCINCAGARELKHLLPOYDQSSFPQSELKPKMTTLVGRLLPVPK 240
QY 61 LNLITQLEGALPSVWNGSAPPSTLPGPKKITLAG 97
DB 241 LNLITQVNDGALPSAVNGAAPPSTLPGPKKITLSG 277

RESULT 12
ID Q9SZD1 PRELIMINARY; PRT; 603 AA.
AC Q9SZD1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Putative transcription factor.
GN F19B15_30 OR AT4G29000 OR AT4G29000/F19B15_30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC SPRAIN-cv, Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078470; CAB43914.1; -.
DR EMBL; AL161574; CAB79638.1; -.
DR EMBL; AK118658; BAC43254.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 603 AA; 64635 MW; 7FD1951AFE818DE5 CRC64;

Query Match 21.7%; Score 357; DB 10; Length 603;
Best Local Similarity 28.2%; Pred. No. 3.7e-28;
Matches 96; Conservative 43; Mismatches 83; Indels 118; Gaps 12;

QY 39 LOSDVP---KP-MTALVGRF-----LPASTKLNLTQOLEGALPSVWNGSAPPSTL 88
DB 58 ISSVSPSTIRPGMTIAGQVTRPTLPMTATMSNPPSQSQ-----IVNA---PIRHPI 109

```

```

QY 89 GPPKI-----TLGYCOCFASGDFCNMCNCCNCCNLLHH 122
DB 110 ESPKARGPRNVEGDGTPOKKKQCNCNKHSLKLYCECFASGYCDGCNCVNCFNVDN 169
QY 123 DIERFKAICACLRNPEAFQPKIGKQGLN-----VKPOHNGCNCRRSGCLKN 171
DB 170 EPARREAVEATLERNPFAFRPKIASSPHGRDKREDIGEVLLGKHNKGCHCKSGCLKK 229
QY 172 YCECYEAQIMCSSICKIGCKNYBESPERKTLMSPN-----YMQ-----TGGLESH 219
DB 230 YCECFQANILCSNCKLDCCKNFEGSERQALFGEHSNHNMAYLQQAANAATAICAVGSSG 289
QY 220 YLP-----PTKESGLPRFSDRR--PSS----- 240
DB 290 FAPSPAPKRRKGQELFNQAIKDSRLSHFQVNNRTGGTSGTSPSPVSRAGGNASSV 349
QY 241 -----CISWEVVEATCACLQAQEEAEKHCCK 268
DB 350 PSKEVYRSLADIITQPHDVRALCSVLTVAGEAAKTSTDK 389

RESULT 13
Q9LW71
ID Q9LW71 PRELIMINARY; PRT; 356 AA.
AC Q9LW71;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genomic DNA, chromosome 3, Pl clone: MSL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB012247; BAB02682.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
SQ SEQUENCE 356 AA; 39796 MW; 4FE0EA29DDDBFF58 CRC64;

Query Match 15.9%; Score 262.5; DB 10; Length 356;
Best Local Similarity 34.8%; Pred. No. 1.2e-18;
Matches 56; Conservative 17; Mismatches 47; Indels 41; Gaps 5;

QY 98 YCDCFASGDFCNMCNCCNCCNLLHHDIERFKAICACLRNPEAFQPKIGKQGLGNVAKPQH 157
DB 79 YCDCFASGVVCTDCDCVCHNNSEKCDAREAMVNLGRNPNFSEK-----ALGSLTDNQ 134
QY 158 -----NKGCCNRRSGCLKNYCEYEAQIMCSSICKIGCKNYBESPERKTLMS 205
DB 135 CKAAPDTKPGLLSRGCKKTRCLKYCECFQANILCSNCKLDCCKNFEGSERQALFGEHSNHNMAYLQQAANAATAICAVGSSG 289
QY 206 MPNYMTQGGLESHYLPPTKFS--GLPRFSDRRPSSCISW 244
DB 187 -----FQPPA-FSAHNSPQVYRRRRDRDELTEW 212

```

RESULT 14

Q9CAV1 PRELIMINARY: PRT; 601 AA.
AC Q9CAV1: 17, Created
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical 66.8 kDa protein.
GN T9J14.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansgore W., Unseid M.,
RA Farmanou B., Vallie G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Deleney M., Bouty M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisine N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Millscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shimp S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana".
RL Nature 408:820-822(2000).
DR EMBL; AC009465; AAG51411.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
KW Hypothetical protein.
SQ SEQUENCE 601 AA; 66786 MW; 10848E970D81E022 CRC64;

Query Match 15.8%; Score 261; DB 10; Length 601;
Best Local Similarity 29.7%; Pred. No. 3.2e-18;
Matches 69; Conservative 25; Mismatches 64; Indels 74; Gaps 10;
QY 33 QDNNYLOSD-----VPKPTAL-----VGRLEPASTKLNLTQLEGAL-PSV---- 75
Db 307 RDSNDLPDSTSIKAPSPQNCCLTSKQDTEILPRTGL---HLNGFVNPVSNGR 363
QY 76 -----VNGSAFPGS-----TLPGPKITLAG----- 97
Db 364 KKKIKDGAFFSTFHYNIEDEFSTPVSTRDLVVSVDKIMEPPERSVEGECFDQLMA 423
QY 98 -----YDCDFASGDFCNN-CNCNCCNNLHDIERFKATKACLGRLPEAFQPKIGKQ--- 149
Db 424 MENRYCEGFCGFCSCQNCFNKPIHEDLVNKSREVIKARNPLAFAPKVVSTDTV 483
QY 150 ----LGNVK----POHNGKNCRRSGCLKNYCEYEAQIMCSCICKICGCKN 193
Db 484 IDLWVENSKTASARHKRCNCRSGCKSKYCECFMGMVGCSSNCRMCCKN 535

RESULT 15

Q9ZS22 PRELIMINARY: PRT; 896 AA.
AC Q9ZS22: 10, Created
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Cysteine-rich polycomb-like protein.
GN CPPL.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20345129; PubMed=10859345;
RA Cvitanich C., Pallisgaard N., Nielsen K.A., Hansen A.C., Larsen K.,
RA Pihakaski-Maunsbach K., Marcker K.A., Jensen E.O.,
RT "CPPL, a novel type DNA-binding protein involved in the expression of a soybean leghemoglobin c3 gene".
RL Proc. Natl. Acad. Sci. U.S.A. 97:8163-8168(2000).
DR EMBL; AJ010165; CAA09028.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;
Query Match 15.1%; Score 248.5; DB 10; Length 896;
Best Local Similarity 40.7%; Pred. No. 1e-16;
Matches 50; Conservative 13; Mismatches 31; Indels 29; Gaps 5;
QY 98 YDCDFASGDFCNN-CNCNCCNNLHDIERFKATKACLGRLPEAFQPKIGKGLGNVKKP 155
Db 487 YDCDFASGDFCNN-CNCNCCNNLHDIERFKATKACLGRLPEAFQPKIGKGLGNVKKP 538
QY 156 -----OHNKNCRRSGCLKNYCEYEAQIMCSCICKICGCKN 196
Db 539 TTDISSHMDENLTTPSSARHKRCNCRSGCKSKYCECFMGMVGCSSNCRMCCKN 598
QY 197 SPE 199
Db 599 KKE 601
Search completed: July 24, 2003, 13:47:44
Job time : 63.9141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:22:37 ; Search time 15.101 Seconds
(without alignments)
931.129 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICOLKGGTQMLCIDNSRT.....GRCLSQLHTFEKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1647	100.0	299	1 MTL5_HUMAN	Q9V415 homo sapien
2	1251	76.0	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
3	109.5	6.6	769	1 ITB2_HUMAN	P05107 homo sapien
4	106.5	6.5	3718	1 LMA5_MOUSE	P01001 mus musculus
5	104	6.3	773	1 MES2_CAEEL	O17514 caenorhabdi
6	103.5	6.3	1786	1 LMB1_MOUSE	P02469 mus musculus
7	103	6.3	769	1 ITB2_PIG	P53714 sus scrofa
8	102	6.2	1790	1 LMB1_DROME	P11046 drosophila
9	101.5	6.2	396	1 TFH2_MOUSE	Q9j1b4 mus musculus
10	101	6.1	1700	1 BAR3_CHITE	Q03376 chironomus
11	101	6.1	3011	1 POLG_HCV1	P26664 h genome po
12	100.5	6.1	760	1 E2_DROME	P42124 drosophila
13	100	6.1	3695	1 LMA5_HUMAN	O15230 homo sapien
14	99	6.0	787	1 ITB6_MOUSE	Q920t9 mus musculus
15	98.5	6.0	3106	1 LMA2_MOUSE	Q60675 mus musculus
16	98.5	6.0	5376	1 ZAN_MOUSE	O88799 mus musculus
17	97.5	5.9	328	1 C170_GIALA	P15799 giardia lam
18	97.5	5.9	643	1 CD93_RAT	O9et61 rattus norv
19	97.5	5.9	969	1 PAC4_HUMAN	P29122 homo sapien
20	97.5	5.9	971	1 RECK_HUMAN	O95980 homo sapien
21	97	5.9	395	1 TFH2_HUMAN	I13888 homo sapien
22	97	5.9	798	1 ITB7_HUMAN	P26010 homo sapien
23	97	5.9	3010	1 POLG_HCVJT	Q00269 h genome po
24	96.5	5.9	290	1 Y070_NPVAC	P41470 autographa
25	96.5	5.9	1576	1 YLK3_CAEEL	P41951 caenorhabdi
26	96.5	5.9	3726	1 TRX_DROME	P20659 drosophila
27	96	5.8	1798	1 ITB1_XENLA	P12606 xenopus lae
28	96	5.8	1808	1 TENA_CHICK	P10039 gallus gall
29	96	5.8	3010	1 POLG_HCVBK	P26663 h genome po
30	95.5	5.8	747	1 E2H1_HUMAN	Q92800 homo sapien
31	95.5	5.8	1801	1 LMB2_RAT	P15800 rattus norv
32	94	5.7	3011	1 POLG_HCVH	P27958 h genome po
33	93	5.6	3110	1 LMA2_HUMAN	P24043 homo sapien

ALIGNMENTS

RESULT 1

ID	MTL5_HUMAN	STANDARD;	PRT;	299 AA.
AC	Q9V415;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	Testis (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).			
GN	MTL5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=99208669; PubMed=10191092;			
RA	Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;			
RT	"A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation.";			
RL	Genomics 57:130-136(1999).			
CC	- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.			
CC	- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U86074; AAD24568.1; ..			
DR	Genew; HGNC:7446; MTL5.			
DR	MIM; 604374; ..			
DR	GO; GO:0005505; F:heavy metal binding activity; TAS.			
DR	GO; GO:0006875; P:metal ion homeostasis; TAS.			
DR	GO; GO:0009412; P:response to heavy metal; TAS.			
DR	GO; GO:0007283; P:spermatogenesis; TAS.			
DR	InterPro; IPR005172; CXC.			
DR	Pfam; PF03638; CXC; 1.			
KW	Spermatogenesis.			
FT	DOMAIN 99 191			
FT	SEQUENCE 299 AA; 33110 MW; 8FG79E140F969C11 CRC64;			

Query Match

Best Local Similarity 100.0%; Score 1647; DB 1; Length 299;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVICOLKGGTQMLCIDNSRTRELKALHLVPOQDQNNYLQSDVPKPMALVGRFLPASTK 60	
Db	1	MVICOLKGGTQMLCIDNSRTRELKALHLVPOQDQNNYLQSDVPKPMALVGRFLPASTK 60	
Qy	61	LNLITQOEGALPSVYVNGSAFPSSGTLPGPPKITLAGYCDPASFDCNCCNCCNNL 120	

```
Db 61 LNLITQLEGALPSVNGSAPSPGSTLPGPKITLAGYCDCAFSGDCFNCCNCCNNL 120
Qy 121 HHDIERFAIKACLRNPEAFQPKIGKQGNVPOHNGKNCRRSGCLKNYCEYEAQI 180
Db 121 HHDIERFAIKACLRNPEAFQPKIGKQGNVPOHNGKNCRRSGCLKNYCEYEAQI 180
Qy 181 MCSSTCKICGCKNYEESPERKTLMSMPNMTGLEGSHYLPPTKFGSLPRFSDRRPSS 240
Db 181 MCSSTCKICGCKNYEESPERKTLMSMPNMTGLEGSHYLPPTKFGSLPRFSDRRPSS 240
Qy 241 CISWEVVEATCACLLOAGEAEKEHCKLAQEMILEEFGRCLSQILHTEPKSKGLME 299
Db 241 CISWEVVEATCACLLOAGEAEKEHCKLAQEMILEEFGRCLSQILHTEPKSKGLME 299

RESULT 2
MTL5_MOUSE
ID MTL5_MOUSE STANDARD; PRT; 295 AA.
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RT early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
CC PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77383; AAD24667.1; -
DR EMBL; U67176; AAD24666.1; -
DR MGD; MGI:1340029; Mtl5.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Spermatogenesis.
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Query Match 76.0%; Score 1251; DB 1; Length 295;
Best Local Similarity 76.3%; Pred. No. 9e-102;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICQLGGTOMLCIDNSRTRELKALHLVPOYQDNVQLSDVPKPTALVGRFLPASTK 60
Db 1 MVICQLGGAGMCLIDNGARELKHLLLPYDDQSSFPQSELPRMTTLVGRLLVPAPK 60
Qy 61 LNLITQLEGALPSVNGSAPSPGSTLPGPKITLAGYCDCAFSGDCFNCCNCCNNL 120
Db 61 LNLITQVNGALPSAVNGAAPPSPGALGGPKITLGSVCDCAFSGDCFNCCS----CNL 116
Qy 121 HHDIERFAIKACLRNPEAFQPKIGKQGNVPOHNGKNCRRSGCLKNYCEYEAQI 180
```

```
Db 117 RHELERFAIKACLRNPEAFQPKMGKRLHSGKNCRRSGCLKNYCEYEAQI 176
Qy 181 MCSSTCKICGCKNYEESPERKTLMSMPNMTGLEGSHYLPPTKFGSLPRFSDRRPSS 240
Db 177 MCSSTCKICGCKNYEESPERKTLMSMPNMTGLEGSHYLPPTKFGSLPRFSDRRPSS 236
Qy 241 CISWEVVEATCACLLOAGEAEKEHCKLAQEMILEEFGRCLSQILHTEPKSKGLME 299
Db 237 CISWEVVEATCACLLOAGEAEKEHCKLAQEMILEEFGRCLSQILHTEPKSKGLKIE 295

RESULT 3
ITB2_HUMAN
ID ITB2_HUMAN STANDARD; PRT; 769 AA.
AC P05107; O16418;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-
DE 1/CR3/p150,95 beta-subunit) (CD18) (Complement receptor C3 beta-
DE subunit).
GN ITGB2 OR CD18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87131080; PubMed=3028646;
RA Kishimoto T.K., O'Connor K., Lee A., Roberts T.M., Springer T.A.;
RT "Cloning of the beta subunit of the leukocyte adhesion proteins:
RT homology to an extracellular matrix receptor defines a novel
RT supergene family.";
RL Cell 48:681-690(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92071153; PubMed=1683838;
RA Weitzman J.B., Wells C.E., Wright A.H., Clark P.A., Law S.K.A.;
RT "The gene organisation of the human beta 2 integrin subunit (CD18).";
RL FEBS Lett. 294:97-103(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=2083799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 9-769 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Spleen;
RX MEDLINE=87246525; PubMed=2954816;
RA Law S.K.A., Gagnon J., Hildreth J.E., Wells C.E., Willis A.C.,
RA Wong A.J.;
RT "The primary structure of the beta-subunit of the cell surface
RT adhesion glycoproteins LFA-1, CR3 and p150,95 and its relationship to
RT the fibronectin receptor.";
RL EMBO J. 6:915-919(1987).
RN [5]
RP SEQUENCE OF 347-355 FROM N.A., AND VARIANTS LAD1 SER-351 AND TRP-586.
RX MEDLINE=92147694; PubMed=1346613;
RA Nelson C., Rabb H., Arnaout M.A.;
RT "Genetic cause of leukocyte adhesion molecule deficiency. Abnormal
```


RT splicing and a missense mutation in a conserved region of CD18 impair
 RL cell surface expression of beta 2 integrins.";
 RN J. Biol. Chem. 267:3351-3357(1992).
 [6]
 RP VARIANT S LAD1 PRO-149 AND ARG-169.
 RX MEDLINE-90293696; PubMed-1694220;
 RA Wardlaw A.J., Hibbs M.L., Stacker S.A., Springer T.A.;
 RT "Distinct mutations in two patients with leukocyte adhesion
 RL deficiency and their functional correlates";
 RN J. Exp. Med. 172:335-345(1990).
 [7]
 RP VARIANT LAD1 ARG-169.
 RX MEDLINE-92324303; PubMed-1352501;
 RA Corbi A., Vará A., Ursa A., Rodriguez M.C.G., Fontan G.,
 RN Sanchez-Madrid F.;
 RT "Molecular basis for a severe case of leukocyte adhesion deficiency.";
 RL Eur. J. Immunol. 22:1877-1881(1992).
 [8]
 RP VARIANT S LAD1 THR-196 AND CYS-593.
 RX MEDLINE-90187099; PubMed-1968911;
 RA Arnsout M.A., Dana N., Gupta S.K., Tenen D.G., Fathallah D.M.;
 RT "Point mutations impairing cell surface expression of the common beta
 RL subunit (CD18) in a patient with leukocyte adhesion molecule
 (leu-CAM) deficiency.";
 RN J. Clin. Invest. 85:977-981(1990).
 [9]
 RP VARIANT LAD1 LEU-178.
 RX MEDLINE-92184805; PubMed-1347532;
 RA Back L.L., Kwok W.W., Hickstein D.D.;
 RT "Identification of two molecular defects in a child with leukocyte
 RL adherence deficiency.";
 RN J. Biol. Chem. 267:5482-5487(1992).
 [10]
 RP VARIANT LAD1 ASN-128.
 RX MEDLINE-92272746; PubMed-1590804;
 RA Matsura S., Kishi F., Tsukahara M., Nunoi H., Matsuda I.,
 RN Kobayashi K., Kajii T.;
 RT "Leukocyte adhesion deficiency: identification of novel mutations in
 RL two Japanese patients with a severe form.";
 RN Biochem. Biophys. Res. Commun. 184:1460-1467(1992).
 [11]
 RP VARIANT LAD1 SER-284.
 RX MEDLINE-93312347; PubMed-7686755;
 RA Back L.A., Kerker M., Baker D., Bauer T.R., Embree L.J.,
 RN Hickstein D.D.;
 RT "A point mutation associated with leukocyte adhesion deficiency type
 RL 1 of moderate severity.";
 RN Biochem. Biophys. Res. Commun. 193:912-918(1993).
 [12]
 RP VARIANT S LAD1 PRO-138 AND ARG-273.
 RX MEDLINE-99102253; PubMed-9884339;
 RA Hogg N., Stewart M.P., Scarth S.L., Newton R., Shaw J.M., Law S.K.A.,
 RN Klein N.;
 RT "A novel leukocyte adhesion deficiency caused by expressed but
 RL nonfunctional beta2 integrins Mac-1 and LFA-1.";
 RN J. Clin. Invest. 103:197-106(1999).
 CC -1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
 CC ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE
 CC RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT
 CC AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE
 CC SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2
 CC RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN
 CC ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-
 CC D/BETA 2 IS A HETERODIMER OF AN ALPHA AND A BETA1.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2
 CC ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.
 CC -1- SUBCELLULAR LOCATION: type I membrane protein.
 CC -1- DISEASE: DEFECTS IN ITGB2 ARE THE CAUSE OF LEUKOCYTE ADHESION
 CC DEFICIENCY, TYPE I (LAD1). LAD1 PATIENTS HAVE RECURRENT BACTERIAL
 CC INFECTIONS AND THEIR LEUKOCYTES ARE DEFICIENT IN A WIDE RANGE OF
 CC ADHESION-DEPENDENT FUNCTIONS.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -1- SIMILARITY: Contains 2 VWFA-like domains.

CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD18 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd18.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y00057; CAA68266.1; -;
 DR EMBL; X64072; CAA45427.1; -;
 DR EMBL; X64073; CAA45427.1; JOINED.
 DR EMBL; X64074; CAA45427.1; JOINED.
 DR EMBL; X64075; CAA45427.1; JOINED.
 DR EMBL; X64076; CAA45427.1; JOINED.
 DR EMBL; X64077; CAA45427.1; JOINED.
 DR EMBL; X64078; CAA45427.1; JOINED.
 DR EMBL; X64079; CAA45427.1; JOINED.
 DR EMBL; X64080; CAA45427.1; JOINED.
 DR EMBL; X64081; CAA45427.1; JOINED.
 DR EMBL; X64082; CAA45427.1; JOINED.
 DR EMBL; X64083; CAA45427.1; JOINED.
 DR EMBL; X63924; CAA45427.1; JOINED.
 DR EMBL; X63925; CAA45427.1; JOINED.
 DR EMBL; X63926; CAA45427.1; JOINED.
 DR EMBL; AL163300; CAB90553.1; -;
 DR EMBL; MI5395; AAA59490.1; -;
 DR EMBL; S81234; AAB21404.1; -;
 DR PIR; A25967; IJHULM.
 DR PDB; 1JX3; 19-SEP-01.
 DR PDB; 1L3Y; 01-APR-02.
 DR Genew; HGNC:6155; ITGB2.
 DR MIM; 600065; -;
 DR MIM; 116920; -;
 DR GO; GO:0008305; C:integrin complex; TAS.
 DR GO; GO:0004895; F:cell adhesion receptor activity; TAS.
 DR GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . . ; TAS.
 DR GO; GO:0007229; P:integrin-mediated signaling pathway; TAS.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR002369; Integrin_B.
 DR InterPro; IPR001169; Integrin_beta_C.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00362; Integrin_B; 1.
 DR PRINTS; PR01186; INTEGRINB.
 DR ProDom; PD001811; Integrin_B; 1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA; 3.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
 KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KW Repeat; Signal; Disease mutation; Pyrrolidone carboxylic acid;
 KW 3D-structure.
 FT SIGNAL 1 22
 FT CHAIN 23 769 INTEGRIN BETA-2
 FT DOMAIN 701 733 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 724 769 POTENTIAL.
 FT DOMAIN 124 363 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 449 617 VWFA-LIKE.
 FT DOMAIN 499 496 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 497 540 I.
 FT REPEAT 541 581 II.
 FT REPEAT 582 617 IV.
 Query Match 6.6%; Score 109.5; DB 1; Length 769;
 Best Local Similarity 22.3%; Pred. No. 0.067; Mismatches 26; Gaps 15;
 Matches 53; Conservative 26; Indels 75;

```

QY 1 MVICOLKGGTQMLCINDSRTELKALHLVPOYQDNNYVLSQDVPKPMYALVGRFLPASTK 60
Db 436 IVTVQVLPQCBRCRDRSL--CH-----GKGFLEGGICRDTGYIGKNCQQTQ 486
QY 61 LNLITQOLEGA-----LPSVVVNSAFPSGSTLPGPKITILAGVCDCEFA----- 103
Db 487 -GRSSQLEGGSCRDNNNSIICSGIGDCVCGQCLCHTSDVFG--KLIYGVQCECDTINCER 543
QY 104 -SGDFCN--CNCNCCNLLHDIERF-----KAICACLGGRNPEAFOPKIGKGQGLGN 152
Db 544 YNGQVCGGGRGLFCGKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VKPQHNKGCNRRSGGLKNCYECYEAQIM-----CSSIC-----KCIGCKNYEESP 198
Db 596 -----CRC-----NVCECHSGYQLPLCQPCGCPSPCGKYISCAECLKEKGP 638

RESULT 4
ID LMA5_MOUSE STANDARD; PRT; 3718 AA.
AC 061001; 09JH06;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LAMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-92 FROM N.A., AND SEQUENCE OF 41-46.
RX MEDLINE=21818471; PubMed=11829758;
RA Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
RT "Complete sequence, recombinant analysis and binding to laminins and
RT sulphated ligands of the N-terminal domains of laminin alpha3B and
RT alpha5 chains.";
RL Biochem. J. 362:213-221(2002).
RN [2]
RP SEQUENCE OF 84-3718 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96081906; PubMed=7499364;
RA Miner J.H., Lewis R.M., Sanes J.R.;
RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
RT expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
RN [3]
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
CC ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
CC KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
CC AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: CONTAINS 1 LAMININ N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE 1.
CC -!- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ IV DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

```

```

CC CC -----THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AJ293593; CAB99255.1; -
CC EMBL; U37501; AAC53430.1; -
CC PIR; T10053; T10053.
CC HSP; P02468; ITLE.
CC MGD; MGI:105382; Lama5.
CC GO; GO:0005604; C:basement membrane; IDA.
CC GO; GO:0030324; P:lung development; IMP.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR001886; LamNT.
CC Pfam; PF00052; laminin_B; 1.
CC Pfam; PF00053; laminin_EGF; 18.
CC Pfam; PF00054; laminin_G; 2.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGF_LAMININ.
CC ProDom; PD002082; Lam_N2; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC SMART; SM00380; EGF_Lam; 20.
CC SMART; SM00281; LamB; 1.
CC SMART; SM00282; LamG; 5.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 19.
CC PROSITE; PS01186; EGF_2; 3.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
CC PROSITE; PS00025; LAM_G_DOMAIN; 5.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL 1 40
CC CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
CC DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
CC DOMAIN 305 363 LAMININ EGF-LIKE 1.
CC DOMAIN 364 433 LAMININ EGF-LIKE 2.
CC DOMAIN 434 479 LAMININ EGF-LIKE 3.
CC DOMAIN 500 546 LAMININ EGF-LIKE 4.
CC DOMAIN 547 592 LAMININ EGF-LIKE 5.
CC DOMAIN 593 637 LAMININ EGF-LIKE 6.
CC DOMAIN 638 682 LAMININ EGF-LIKE 7.
CC DOMAIN 683 728 LAMININ EGF-LIKE 8.
CC DOMAIN 729 781 LAMININ EGF-LIKE 9.
CC DOMAIN 782 833 LAMININ EGF-LIKE 10.
CC DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
CC DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
CC DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
CC DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
CC DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
CC DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
CC DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
CC DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
CC DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
CC DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
CC DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
CC DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
CC DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
CC DOMAIN 2169 2735 DOMAIN II AND I.
CC DOMAIN 2736 2933 LAMININ G-LIKE 1.
CC DOMAIN 2947 3119 LAMININ G-LIKE 2.
CC DOMAIN 3128 3296 LAMININ G-LIKE 3.
CC DOMAIN 3337 3511 LAMININ G-LIKE 4.
CC DOMAIN 3518 3689 LAMININ G-LIKE 5.
CC DOMAIN 2205 2257 COILED COIL (POTENTIAL).

```


DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001886; LamNT.
 DR Pfam: PF00053; laminin_EGF; 13.
 DR Pfam: PF00055; laminin_Nterm; 1.
 DR PRINTS: PRO0011; EGFLAMININ.
 DR SMART: SM00180; EGF_Lam; 11.
 DR SMART: SM00136; LamNT; 1.
 DR PROSITE: PS00022; EGF_1; 9.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 1786
 FT DOMAIN 22 270 LAMININ BETA-1 CHAIN.
 FT DOMAIN 271 334 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 335 337 LAMININ EGF-LIKE 1.
 FT DOMAIN 338 457 LAMININ EGF-LIKE 2.
 FT DOMAIN 458 509 LAMININ EGF-LIKE 3.
 FT DOMAIN 510 540 LAMININ EGF-LIKE 4.
 FT DOMAIN 541 772 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 773 820 LAMININ DOMAIN IV.
 FT DOMAIN 821 866 LAMININ EGF-LIKE 6.
 FT DOMAIN 867 916 LAMININ EGF-LIKE 7.
 FT DOMAIN 917 975 LAMININ EGF-LIKE 8.
 FT DOMAIN 976 1027 LAMININ EGF-LIKE 9.
 FT DOMAIN 1028 1083 LAMININ EGF-LIKE 10.
 FT DOMAIN 1084 1131 LAMININ EGF-LIKE 11.
 FT DOMAIN 1132 1178 LAMININ EGF-LIKE 12.
 FT DOMAIN 1179 1397 LAMININ EGF-LIKE 13.
 FT DOMAIN 1398 1430 DOMAIN II.
 FT DOMAIN 1431 1786 DOMAIN ALPHA.
 FT DOMAIN 1787 1831 COILED COIL (POTENTIAL).
 FT DOMAIN 1832 1881 COILED COIL (POTENTIAL).
 FT DOMAIN 1882 1931 COILED COIL (POTENTIAL).
 FT DOMAIN 1932 2080 BY SIMILARITY.
 FT DISULFID 271 280
 FT DISULFID 273 298
 FT DISULFID 300 309
 FT DISULFID 312 332
 FT DISULFID 332 332
 FT DISULFID 335 344
 FT DISULFID 337 362
 FT DISULFID 365 374
 FT DISULFID 377 395
 FT DISULFID 398 411
 FT DISULFID 400 426
 FT DISULFID 428 437
 FT DISULFID 440 455
 FT DISULFID 458 472
 FT DISULFID 460 479
 FT DISULFID 481 490
 FT DISULFID 493 507
 FT DISULFID 507 572
 FT DISULFID 773 785
 FT DISULFID 775 792
 FT DISULFID 794 803
 FT DISULFID 806 818
 FT DISULFID 821 833
 FT DISULFID 823 840
 FT DISULFID 842 851
 FT DISULFID 854 864
 FT DISULFID 867 876
 FT DISULFID 869 883
 FT DISULFID 886 895
 FT DISULFID 898 914
 FT DISULFID 917 933
 FT DISULFID 919 944
 FT DISULFID 946 955
 FT DISULFID 958 973
 FT DISULFID 976 990
 FT DISULFID 978 997
 FT DISULFID 1000 1009
 FT DISULFID 1012 1025
 FT DISULFID 1084 1096
 FT DISULFID 1086 1103

FT DISULFID 1105 1114 BY SIMILARITY.
 FT DISULFID 1117 1129 BY SIMILARITY.
 FT DISULFID 1132 1144 BY SIMILARITY.
 FT DISULFID 1134 1151 BY SIMILARITY.
 FT DISULFID 1153 1162 BY SIMILARITY.
 FT DISULFID 1165 1176 BY SIMILARITY.
 FT DISULFID 1179 1182 INTERCHAIN (PROBABLE).
 FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).
 FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 577 677 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1041 1041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1195 1195 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1279 1279 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1343 1343 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1533 1533 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1542 1542 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1643 1643 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 1531 1534 SGNA -> NEMP (IN REF. 2).
 FT CONFLICT 1749 1749 D -> N (IN REF. 2).
 SQ SEQUENCE 1786 AA; 196904 MW; 846671B7BF41A74 CRC64;
 Query Match 6.3%; Score 103.5; DB 1; Length 1786;
 Best Local Similarity 21.4%; Pred. No. 0.58; Indels 95; Gaps 15;
 Matches 54; Conservative 29; Mismatches 74; Indels 95; Gaps 15;
 QY 90 PPKITLAGYCDGFCAGDFFCNC-----NCNNCCNHLHDIERRKAIKACLR 136
 Db 937 PVTQLACVCDPGYIGSRCDDCAGFFGNPSDFGSCQPC--QCHNIDT-----T 985
 QY 137 NPEAFQPKIGK-----GQIGNVKVQHNKGCNCRSRGCLKNY-----C 173
 Db 986 DPEACDKDTGCLKLYHTEGDHCQLCQGYGYGDALRQDCRCVCVNYLGTVEHCNGSDC 1045
 QY 174 ECYEAQIMCSSICKIGCKNYEESPERKTLMSNPWMO---TG---GLESHYLPPT- 224
 Db 1046 HCDKATGQCCLPNVIG-----QNCDR-----CAPNTWQLASGTGCGPCNCAHSGPSC 1096
 QY 225 -KFSG---LPRFSDHRRPSSC--ISW-----EYVEATCAGLLAQ 257
 Db 1097 NEFTGOCOCMPGFG-GRTCECQELFWGDPDVECRACDPRGIETPQCDQSTGQCVCVE 1155
 QY 258 GEEAEKEHCKSC 269
 Db 1156 GVEGPR--CDKC 1165
 RESULT 7
 ITB2_PIG STANDARD; PRT; 769 AA.
 AC P53714;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-
 DE 1/CR3/PL150.95 beta-subunit) (CD18) (Complement receptor C3 beta-
 DE subunit).
 GN ITGB2 OR CD18.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RP [1]
 RN SEQUENCE FROM N.A.
 RA Lee J.K., Schook L.B., Rutherford M.S.;
 RT "Molecular cloning and characterization of the porcine CD18 leukocyte
 RT adhesion molecule.";
 RL Xenotransplantation 3:222-230(1996).
 CC -!- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,

CC ICM3 AND ICM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE
 CC RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT
 CC AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE
 CC SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2
 CC RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN
 CC ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-
 CC D/BETA-2 IS A RECEPTOR FOR ICM3 AND VCAM1.
 CC SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2
 CC ASSOCIATES WITH EITHER ALPHA-1, ALPHA-M, ALPHA-X OR ALPHA-D.
 CC SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -!- SIMILARITY: Contains 2 WFA-like domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U13941; AAB16868.1; -.
 CC HSP: P05106; 1JY2.
 CC InterPro: IPR006209; EGF-like.
 CC InterPro: IPR002369; Integrin_B.
 CC InterPro: IPR001169; Integrin_Beta_C.
 CC InterPro: IPR003659; Plexin-like.
 CC InterPro: IPR002035; WFA.
 CC Pfam: PF00362; Integrin_B; 1.
 CC PRINTS: PR01186; INTEGRINB.
 CC ProDom: PR001811; Integrin_B; 1.
 CC SMART: SM00187; INB; 1.
 CC SMART: SM00423; PSI; 1.
 CC SMART: SM00327; WFA; 1.
 CC PROSITE: PS00243; INTEGRIN_BETA; 3.
 CC PROSITE: PS00022; EGF_1; UNKNOWN_2.
 CC PROSITE: PS01186; EGF_2; UNKNOWN_3.
 KW Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 769
 FT DOMAIN 23 700
 FT TRANSMEM 701 723
 FT DOMAIN 724 769
 FT DOMAIN 124 363
 FT DOMAIN 449 617
 FT REPEAT 449 496
 FT REPEAT 497 540
 FT REPEAT 541 581
 FT REPEAT 582 617
 FT DISULFID 25 447
 FT DISULFID 33 43
 FT DISULFID 36 73
 FT DISULFID 46 62
 FT DISULFID 191 198
 FT DISULFID 246 286
 FT DISULFID 386 400
 FT DISULFID 420 662
 FT DISULFID 445 449
 FT DISULFID 459 470
 FT DISULFID 467 506
 FT DISULFID 472 481
 FT DISULFID 483 497
 FT DISULFID 512 517
 FT DISULFID 514 549
 FT DISULFID 519 534
 FT DISULFID 536 541
 FT DISULFID 557 562
 FT DISULFID 559 590
 FT DISULFID 564 573
 FT DISULFID 575 582
 FT DISULFID 596 601
 FT DISULFID 598 643
 FT DISULFID 603 612

FT DISULFID 615 618
 FT DISULFID 622 631
 FT DISULFID 628 695
 FT DISULFID 647 670
 FT SITE 397 399
 FT CARBOHYD 50 50
 FT CARBOHYD 116 116
 FT CARBOHYD 254 254
 FT CARBOHYD 501 501
 FT CARBOHYD 642 642
 SQ SEQUENCE 769 AA; 84789 MW; FDD606CEE850449 CRC64;
 Query Match 6.3%; Score 103; DB 1; Length 769;
 Best Local Similarity 21.5%; Pred. NO. 0.25;
 Matches 59; Conservative 28; Mismatches 88; Indels 100; Gaps 16;
 Qy 14 CIDSRTRFKALHLVPOYQDNNYVLOSVDVPPKMTALVGRPLASTKLNLTQOEGALP 73
 Db 449 CGDSKERTLCG-----NKGSMGCVGRCDAGYIGKHCECQTQ-GRSSQLEGSCR 498
 Qy 74 SVVNGSAFFPGSTLPGPKITLAGYDCFCAS-----GDFCNMNCNCCNLLHHDIE 125
 Db 499 K-DNSSIITCSGL-----GDCICGCVCHTSDVPNKIYQFC-ECDNMNC-----E 542
 Qy 126 RFKAIKACLG-----RNPEAPQPKIGKGLGNVKKFOHNGKNCRRS--GCLK-- 170
 Db 543 RFDG-QVCGGKRGKLCFCSTCRQEGFE-----GSACQCLKSTQGLNLQ 586
 Qy 171 -----NYCEC--YE-----AQIMSSICKICGCKNYEESPERKTLMSM 206
 Db 587 GVECSGRGRRCNVCCDFGYPPLCTDCPCQVPCARYAKACELKFDTPFFAKNSAE 646
 Qy 207 PNYMTQGLESHYLPPTKFSGLPRFSDRRPSSC 241
 Db 647 C-----GTTKLLPSRMGS--RKCNERDSEGC 670
 RESULT 8
 LMBI_DROME STANDARD; PRT; 1790 AA.
 ID LMBI_DROME STANDARD; Q9VLMW6; Q9XZT4;
 AC P11046; Q26328; Q9VLMW6; Q9XZT4;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Laminin beta-1 chain precursor (Laminin BI chain).
 GN LANBI OR LAMB1 OR CG7123.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=94000382; PubMed=8397815;
 RA Gow C.-H., Chang H.-Y., Lin C.-J., Chiang T.-W., Hui C.-F.;
 RT "Analysis of the Drosophila gene for the laminin BI chain";
 RL DNA Cell Biol. 12:573-587(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88210471; PubMed=3365769;
 RA Montell D.J., Goodman C.S.;
 RT "Drosophila substrate adhesion molecule: sequence of laminin BI chain
 reveals domains of homology with mouse";
 RL Cell 53:463-473(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwaku C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 CC is thought to mediate the attachment, migration, and organization
 CC of cells into tissues during embryonic development by interacting
 CC with other extracellular matrix components.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -!- SIMILARITY: Contains 13 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 2 laminin IV domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M95811; AAD19752.1; -;
 CC EMBL: M19525; AAA28663.1; -;
 CC EMBL: AF003618; AAF52563.1; -;
 CC PIR: A28783; MWFFS1.
 CC HSP: P02468; 1TLE.
 CC FlyBase: FBgn0002527; LanB1.
 CC GO: GO:0005605; C:basal lamina; IDA.
 CC InterPro: IPR0006209; EGF_like.
 CC InterPro: IPR002049; Laminin_EGF.
 CC InterPro: IPR001886; LamNT.
 CC Pfam: PF00053; laminin_EGF; 13.
 CC Pfam: PF00055; laminin_Nterm; 1.

DR PRINTS; PRO0011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 12.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 10.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1790 LAMININ BETA-1 CHAIN.
 FT DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 290 356 LAMININ EGF-LIKE 1.
 FT DOMAIN 337 419 LAMININ EGF-LIKE 2.
 FT DOMAIN 420 479 LAMININ EGF-LIKE 3.
 FT DOMAIN 480 530 LAMININ EGF-LIKE 4.
 FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 562 789 LAMININ DOMAIN IV.
 FT DOMAIN 791 838 LAMININ EGF-LIKE 6.
 FT DOMAIN 839 884 LAMININ EGF-LIKE 7.
 FT DOMAIN 885 934 LAMININ EGF-LIKE 8.
 FT DOMAIN 935 992 LAMININ EGF-LIKE 9.
 FT DOMAIN 993 1044 LAMININ EGF-LIKE 10.
 FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.
 FT DOMAIN 1096 1143 LAMININ EGF-LIKE 12.
 FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
 FT DOMAIN 1191 1407 DOMAIN II.
 FT DOMAIN 1408 1434 DOMAIN ALPHA.
 FT DOMAIN 1435 1790 DOMAIN I.
 FT DOMAIN 1257 1470 COILED COIL (POTENTIAL).
 FT DOMAIN 1455 1507 COILED COIL (POTENTIAL).
 FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).
 FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).
 FT DISULFID 290 320 BY SIMILARITY.
 FT DISULFID 322 331 BY SIMILARITY.
 FT DISULFID 334 354 BY SIMILARITY.
 FT DISULFID 357 366 BY SIMILARITY.
 FT DISULFID 359 384 BY SIMILARITY.
 FT DISULFID 387 396 BY SIMILARITY.
 FT DISULFID 399 417 BY SIMILARITY.
 FT DISULFID 420 433 BY SIMILARITY.
 FT DISULFID 422 448 BY SIMILARITY.
 FT DISULFID 450 459 BY SIMILARITY.
 FT DISULFID 462 477 BY SIMILARITY.
 FT DISULFID 480 493 BY SIMILARITY.
 FT DISULFID 482 500 BY SIMILARITY.
 FT DISULFID 502 511 BY SIMILARITY.
 FT DISULFID 514 528 BY SIMILARITY.
 FT DISULFID 791 803 BY SIMILARITY.
 FT DISULFID 812 821 BY SIMILARITY.
 FT DISULFID 824 836 BY SIMILARITY.
 FT DISULFID 839 851 BY SIMILARITY.
 FT DISULFID 841 858 BY SIMILARITY.
 FT DISULFID 860 869 BY SIMILARITY.
 FT DISULFID 872 882 BY SIMILARITY.
 FT DISULFID 885 894 BY SIMILARITY.
 FT DISULFID 887 901 BY SIMILARITY.
 FT DISULFID 904 913 BY SIMILARITY.
 FT DISULFID 916 932 BY SIMILARITY.
 FT DISULFID 935 951 BY SIMILARITY.
 FT DISULFID 937 962 BY SIMILARITY.
 FT DISULFID 964 973 BY SIMILARITY.
 FT DISULFID 976 990 BY SIMILARITY.
 FT DISULFID 993 1007 BY SIMILARITY.
 FT DISULFID 995 1014 BY SIMILARITY.
 FT DISULFID 1017 1026 BY SIMILARITY.
 FT DISULFID 1029 1042 BY SIMILARITY.
 FT DISULFID 1045 1059 BY SIMILARITY.
 FT DISULFID 1047 1066 BY SIMILARITY.
 FT DISULFID 1068 1077 BY SIMILARITY.
 FT DISULFID 1080 1093 BY SIMILARITY.
 FT DISULFID 1096 1108 BY SIMILARITY.

```

FT DISULFID 1098 1115 BY SIMILARITY.
FT DISULFID 1117 1126 BY SIMILARITY.
FT DISULFID 1129 1141 BY SIMILARITY.
FT DISULFID 1144 1156 BY SIMILARITY.
FT DISULFID 1146 1163 BY SIMILARITY.
FT DISULFID 1165 1174 BY SIMILARITY.
FT DISULFID 1177 1188 BY SIMILARITY.
FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1788 1788 INTERCHAIN (PROBABLE).
FT SITE 643 645 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.2%; Score 102; DB 1; Length 1790;
Best Local Similarity 21.4%; Pred. No. 0.79;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

Qy 97 GYCD-----CF--ASGDFCNNC-----NCNNC-----CNLHHDIERFKA 129
Db 1005 GNCDRQTGACLKCLYQTTGDHCELCCKDFFGDLQNCQCCDFLGTNTNTIAHCDRTG 1064

Qy 130 IKACLGRLNPEAFQPKIGKGLGNVKKPOHKN-----GCNRRSGCLKNYCEYEAQIM 181
Db 1065 QCPCL-----ENVQGVRCDOCAENHWKIASGEGCEGCNCDPIGALHEOCNSYTGQCQ 1116

Qy 182 CS---SICKIGCK-NVEESPRTKLSMPNMTQGLGSHLYLPPTKFS--GLPFSHD 235
Db 1117 CKPGFGRACNQCQAHWTGNPEK-----CQCECQDFGAADFQCD 1157

Qy 236 RPPSSCISWEVEATCACLAAQGEABEKHCSCLAQEMILEEFGRLS 284
Db 1158 RETGNCVCHGIGGYKNCNARGIYGFPHCSPC-----GECFN 1196

RESULT 9
TFH2_MOUSE STANDARD; PRT; 396 AA.
ID TFH2_MOUSE
AC Q9JIB4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE TFIIH basal transcription factor complex p44 subunit (Basic
DE transcription factor 2.44 kDa subunit) (BTF2-p44) (General
DE transcription factor ITH polypeptide 2).
GN GTF2H2 OR BTF2P44.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2041147; PubMed=10958627;
RA Endrizzi M.G., Hadinoto V., Gowney J.D., Miller W., Dietrich W.F.;
RT "Genomic sequence analysis of the mouse Naip gene array.";
RL Genome Res. 10:1095-1102(2000).
CC -1- FUNCTION: Component of the core-TFIIH basal transcription factor
CC involved in nucleotide excision repair (NER) of DNA and, when
CC complexed to CAK, in RNA transcription by RNA polymerase II. The
CC N-terminus interacts with and regulates XPD whereas an intact C-
CC terminus is required for a successful escape of RNAP II form the
CC promoter (By similarity).
CC -1- SUBUNIT: One of the six subunits forming the core-TFIIH basal
CC transcription factor. Interacts with XPD, XPD, p62 and p34 (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.

```

```

CC -1- SIMILARITY: Contains 1 C4-type zinc finger.
CC -1- SIMILARITY: Contains 1 VMFA domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF242432; AAF82753.1;
CC MGI; MGI:1345669; Gtf2h2.
CC InterPro; IPR004595; Ssl1.
CC Pfam; PF04056; Ssl1; 1.
CC TIGRFAMs; TIGR00622; ssl1; 1.
CC PROSITE; PS0234; VMFA; 1.
CC Transcription regulation; DNA repair; Nuclear protein; Zinc-finger.
KW DOMAIN 60 236 VMFA.
FT ZN_FING 292 309 C4-TYPE.
FT DOMAIN 242 245 POLY-SER.
SQ SEQUENCE 396 AA; 44687 MW; 1BC57AFD464AA83F CRC64;

Query Match 6.2%; Score 101.5; DB 1; Length 396;
Best Local Similarity 21.8%; Pred. No. 0.16;
Matches 48; Conservative 27; Mismatches 72; Indels 73; Gaps 11;

Qy 2 VICOLKGTOMLCIDNSTRRELKALHLVPOYQDONNVYLSQSDVPKPMYALVGRFLPASTKL 61
Db 210 VLARETGGTVHVLDETHYKELLAHVSPPPASSSS-----ECSLRMGFPQHTIA 260

Qy 62 NLITQOLEGALPVVNGSAFPGSTLPGPKKITLAGYCDGCFASGDGFCNCCNCC----- 116
Db 261 SLSDD---AKPS-FSAHLDDNNSTEG---LILGGY-----FCPCRAKYCELPVE 305

Qy 117 -----CNLHH-----DIERFKAICAKLGNRNPFAOPKIGKGLGN 152
Db 306 CKICGLTIVSAPHLARSYVHPLDADFQETISLEEKYKGERFCYCGQ-----GE 352

Qy 153 VKQHNKGCN-CRRSGCLKNYCEYEAQIMCSSICKICIGC 191
Db 353 LKQDHVYVYTCQNVFCVD--CDVF-----VHDSLHCPCGC 386

RESULT 10
BAR3_CHITE STANDARD; PRT; 1700 AA.
ID BAR3_CHITE
AC Q03376;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE Balbiani ring protein 3 precursor.
DE Balbiani ring protein 3 precursor.
GN BR3.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomidae;
OC Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90172404; PubMed=1689777;
RA Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
RT repetitive structure split by many introns.";
RL J. Mol. Biol. 211:331-349(1990).
CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
CC -1- SUBCELLULAR LOCATION: Secreted.

```


CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.
 CC -1- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X52263; CAJ36506.1; -
 CC PIR; S08167; S08167.
 CC HSP; P15358; 1SKZ.
 CC InterPro; IPR004153; CXKXC_repeat.
 CC Pfam; PF03128; CXKXC; 71.
 KW Repeat; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.
 SQ SEQUENCE 1700 AA; 186145 MW; 34202828521B0815 CRC64;
 Query Match 6.1%; Score 101; DB 1; Length 1700;
 Best Local Similarity 22.6%; Pred. No. 0.91;
 Matches 51; Conservative 20; Mismatches 87; Indels 68; Gaps 15;
 QY 99 CDCFASGD-----FC-NNCNC-----NNCCNLHDIERFKAICACLRNPEAF 141
 DB 198 CSCECKGDKGCGSKWCKNCRICPTAEPAGGSAFLKWD--KSCACPAKWEKK 255
 QY 142 QPKIGKGLGNVYKPHNKGNCRRSGCLKN-----YCECYE-----AQIMCSSICK 187
 DB 256 EKCVESGKIWN---PNTCECGAQLNCPDNKANKETCOCECKEVKCKNGGVFCCKSCS 312
 QY 188 CIGCKNYEESPERRKTLMSPNYMTQGLGSHYLP---PTKPSGLPR----- 231
 DB 313 CV-CFPGDKD---KCTAPQVY---DGVACSCSPVNMQKPADGCPKQKWDKECRCEC 365
 QY 232 -FSHRRPSSCISWEVATAC-----LLAQGEAEKEHCS-KCL 270
 DB 366 PVKHCKNGKV-WDETICQICPRDAPVCTAGKRGESCECKCI 409
 RESULT 11
 POLG_HCV1
 ID POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OC NCBI_TaxID=11104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RL "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M62321; AAA45676.1; -
 CC PIR; A39166; GNMVCS.
 CC PDB; 1A1V; 16-FEB-99.
 CC PDB; 1HEI; 25-NOV-98.
 CC MEROPS; S29.001; -
 CC MEROPS; U39.001; -
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4a.
 CC InterPro; IPR001490; HCV_NS4b.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RDRP.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR007095; RNA_pol_DS_Ps.
 CC InterPro; IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.
 CC Pfam; PF01542; HCV_core; 1.
 CC Pfam; PF01539; HCV_env; 1.
 CC Pfam; PF01560; HCV_NS1; 1.
 CC Pfam; PF01538; HCV_NS2; 1.
 CC Pfam; PF02907; HCV_NS3; 1.
 CC Pfam; PF01006; HCV_NS4a; 1.
 CC Pfam; PF01001; HCV_NS4b; 1.
 CC Pfam; PF01506; HCV_NS5a; 1.
 CC Pfam; PF00271; helicase_C; 1.
 CC Pfam; PF00998; Viral_RDRP; 1.
 CC ProDom; PD186062; HCV_NS1; 1.
 CC SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3012 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 369 1083 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC...) (POTENTIAL).

```

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 3011;
Best Local Similarity 24.8%; Pred. No. 1.8;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

Oy 79 SAFP-SGSTLPGPKITLAGYCDGFCAGDCNCCNCCNNL-----HHDIERPKA1 130
| | | | | : | | | | | : | | | | | : | | | | |
Db 1424 SVIPTSDDVVVATDALMTGY-----TGDFDSVIDCNCVTQTVDLSLDTFTTITL 1478
| | | | | : | | | | | : | | | | | : | | | | |

Oy 131 KACLGIRPEAFOPKIGKGLG---NVKPKHKGKNCRRSGCLKN--YCECYEAQIMCSSI 185
| | | | | : | | | | | : | | | | | : | | | | |
Db 1479 QDAVSRTOR---RGRTRGKPGIYRFVAPGE-----RPSGMFDSVLCCECYDA----- 1523

Oy 186 CKICGCKNYEBSPERKTLMSNPYMTQGGT-----EG-----SHVLPPTKF 226
| | | | | : | | | | | : | | | | | : | | | | |
Db 1524 ---GCAYWELTP-AETVRLRAYMTPGLPVQCDHLEFWGVTGLTHDAHLSQTK 1578

Oy 227 SG--LPRFSDRRPSSCISWEVBEATCACLIAQGEAEKHCSCKL 270
| | | | | : | | | | | : | | | | | : | | | | |
Db 1579 SGENLP-----YLVAYQATVCARAQAAPPSPDQMWKCL 1611

RESULT 12
EZ_DROME STANDARD; PRT; 760 AA.
AC P42124; OSVTA3;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polycomb protein E(z) (Enhancer of zeste protein).
GN E(z) OR CG6502.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019309; PubMed=8413234;
RA Jones R.S., Gelbart W.M.;
RT "The Drosophila Polycomb-group gene Enhancer of zeste contains a
region with sequence similarity to trithorax."
RL Mol. Cell. Biol. 13:6357-6366(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley.
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Brandon K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

```

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flisler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley.
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=97164720; PubMed=9012527;
RA Carrington E.A., Jones R.S.;
RT "The Drosophila Enhancer of zeste gene encodes a chromosomal protein:
examination of wild-type and mutant protein distribution."
RL Development 122:4073-4083(1996).
RN [5]
RP INTERACTION WITH ESC.
RX MEDLINE=98226554; PubMed=9566901;
RA Jones C.A., Ng J., Peterson A.J., Morgan K., Simon J.A., Jones R.S.;
RT "The Drosophila esc and E(z) proteins are direct partners in polycomb
group-mediated repression."
RL Mol. Cell. Biol. 18:2825-2834(1998).
RN [6]
RP IDENTIFICATION IN A COMPLEX WITH RPD3; ESC AND CAP1.
RX MEDLINE=21064443; PubMed=11124122;
RA Tie F., Furuyama T., Prasad-Sinha J., Jane E., Harte P.J.;
RT "The Drosophila Polycomb group proteins ESC and E(z) are present in a
complex containing the histone-binding protein p55 and the histone
deacetylase RPD3."
RL Development 128:275-286(2001).
RN [7]
RP IDENTIFICATION IN A COMPLEX WITH RPD3; PHO AND ESC, AND
TRANSIENT INTERACTION WITH THE PRC1 COMPLEX.
RX MEDLINE=21464651; PubMed=11581156;
RA Poux S., Melfi R., Pirrotta V.;
RT "Establishment of Polycomb silencing requires a transient interaction
between PC and ESC."
RL Genes Dev. 15:2509-2514(2001).
RN [8]
RP IDENTIFICATION IN A ESC/E(z) COMPLEX WITH ESC; RPD3; CAP1 AND SU(2)12,
METHYLTRANSFERASE ACTIVITY OF THE COMPLEX, AND MUTAGENESIS OF

```

[illegible]

RX MEDLINE-21456161; PubMed-11572484;
 RA Nagase T., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XXI.
 RT The complete sequences of 60 new cDNA clones from brain which code for
 RT large proteins.";
 RL DNA Res. 8:179-187(2001).
 RN [3]
 RP SEQUENCE OF 2051-3695 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE-98290545; PubMed-9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 RN [4]
 RP SEQUENCE OF 2743-3695 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE-97415425; PubMed-9271224;
 RA Durkin M.E., Loechel F., Mattel M.-G., Gilpin B.J., Albrechtsen R.,
 RA Wever U.M.;
 RT "Tissue-specific expression of the human laminin alpha5-chain, and
 RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal
 RT mouse chromosome 2 near the locus for the ragged (Ra) mutation.";
 RL FEBS Lett. 411:296-300(1997).
 RN [5]
 RP EXPRESSION IN RETINA.
 RX MEDLINE-20422761; PubMed-10964957;
 RA Libby R.T., Champliand M.-F., Claudepierre T., Xu Y., Gibbons E.P.,
 RA Koch M., Burgeson R.E., Hunter D.D., Brunken W.J.;
 RT "Laminin expression in adult and developing retinae: evidence of two
 RT novel CNS laminins.";
 RL J. Neurosci. 20:6517-6528(2000).
 CC -|- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION AND ORGANIZATION
 CC OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -|- SUBUNIT: LAMININ-15 COMPLEX IS AN HETEROTRIMER COMPOSED OF THREE
 CC CHAINS (ALPHA-5/BETA-2/GAMMA-3) WHICH ARE BOUND TO EACH OTHER BY
 CC DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG
 CC AND THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC -|- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 CC MEMBRANES (MAJOR COMPONENT).
 CC -|- TISSUE SPECIFICITY: EXPRESSED IN HEART, LUNG, KIDNEY, SKELETAL
 CC MUSCLE, PANCREAS, RETINA AND PLACENTA. LITTLE OR NO EXPRESSION IN
 CC BRAIN AND LIVER.
 CC -|- DOMAIN: DOMAIN G IS GLOBULAR AND IS PART OF THE MAJOR CELL-BINDING
 CC SITE LOCATED IN THE LONG ARM OF THE LAMININ HETEROTRIMER.
 CC -|- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -|- SIMILARITY: Contains 22 laminin EGF-like domains.
 CC -|- SIMILARITY: Contains 2 laminin IV domains.
 CC -|- SIMILARITY: Contains 5 laminin G-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL354836; CAC22309.1; ALT_SEQ.
 DR EMBL; AL354836; CAC22310.1; -.
 DR EMBL; AB067494; BAB67800.1; -.
 DR EMBL; AB011105; BAA25459.1; -.
 DR EMBL; Z95636; CAB09137.1; -.
 DR HSP; P02468; IKLO.
 DR Genew; HGNC:6485; LAMA5.
 DR MIM; 601033; -.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.

DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR001886; LamNT.
 DR Pfam; PF00052; laminin_B; 1.
 DR Pfam; PF00053; laminin_EGF; 18.
 DR Pfam; PF00054; laminin_G; 2.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR ProDom; PD002082; Lam_N2; 1.
 DR ProDom; PD003031; Laminin_B; 1.
 DR SMART; SM00180; EGF_Lam; 20.
 DR SMART; SM00281; LamB; 1.
 DR SMART; SM00282; LamC; 5.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 35
 FT CHAIN 36 3695 LAMININ ALPHA-5 CHAIN.
 FT DOMAIN 36 299 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 300 358 LAMININ EGF-LIKE 1.
 FT DOMAIN 359 428 LAMININ EGF-LIKE 2.
 FT DOMAIN 429 474 LAMININ EGF-LIKE 3.
 FT DOMAIN 494 540 LAMININ EGF-LIKE 4.
 FT DOMAIN 541 586 LAMININ EGF-LIKE 5.
 FT DOMAIN 587 631 LAMININ EGF-LIKE 6.
 FT DOMAIN 632 676 LAMININ EGF-LIKE 7.
 FT DOMAIN 677 722 LAMININ EGF-LIKE 8.
 FT DOMAIN 723 775 LAMININ EGF-LIKE 9.
 FT DOMAIN 776 828 LAMININ EGF-LIKE 10.
 FT DOMAIN 829 850 LAMININ EGF-LIKE 11 (INCOMPLETE).
 FT DOMAIN 851 1437 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.
 FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.
 FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.
 FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.
 FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).
 FT DOMAIN 1638 1830 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1831 1863 LAMININ EGF-LIKE 16 (C-TERMINAL).
 FT DOMAIN 1864 1912 LAMININ EGF-LIKE 17.
 FT DOMAIN 1913 1968 LAMININ EGF-LIKE 18.
 FT DOMAIN 1969 2022 LAMININ EGF-LIKE 19.
 FT DOMAIN 2023 2069 LAMININ EGF-LIKE 20.
 FT DOMAIN 2070 2116 LAMININ EGF-LIKE 21.
 FT DOMAIN 2117 2166 LAMININ EGF-LIKE 22.
 FT DOMAIN 2167 2735 DOMAIN II AND I.
 FT DOMAIN 2736 2929 LAMININ G-LIKE 1.
 FT DOMAIN 2941 3115 LAMININ G-LIKE 2.
 FT DOMAIN 3124 3292 LAMININ G-LIKE 3.
 FT DOMAIN 3340 3513 LAMININ G-LIKE 4.
 FT DOMAIN 3520 3692 LAMININ G-LIKE 5.
 FT DOMAIN 2203 2221 COILED COIL (POTENTIAL).
 FT DOMAIN 2335 2466 COILED COIL (POTENTIAL).
 FT DOMAIN 2510 2570 COILED COIL (POTENTIAL).
 FT SITE 1722 1724 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1838 1840 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 300 309 BY SIMILARITY.
 FT DISULFID 302 322 BY SIMILARITY.
 FT DISULFID 324 333 BY SIMILARITY.
 FT DISULFID 336 356 BY SIMILARITY.
 FT DISULFID 359 368 BY SIMILARITY.
 FT DISULFID 361 393 BY SIMILARITY.
 FT DISULFID 396 405 BY SIMILARITY.
 FT DISULFID 408 426 BY SIMILARITY.
 FT DISULFID 429 440 BY SIMILARITY.
 FT DISULFID 431 447 BY SIMILARITY.
 FT DISULFID 449 458 BY SIMILARITY.
 FT DISULFID 461 471 BY SIMILARITY.
 FT DISULFID 494 506 BY SIMILARITY.
 FT DISULFID 496 515 BY SIMILARITY.
 FT DISULFID 517 526 BY SIMILARITY.

FT DISULFID 529 538 BY SIMILARITY.
 FT DISULFID 541 553 BY SIMILARITY.
 FT DISULFID 543 560 BY SIMILARITY.
 FT DISULFID 562 571 BY SIMILARITY.
 FT DISULFID 574 584 BY SIMILARITY.
 FT DISULFID 587 599 BY SIMILARITY.
 FT DISULFID 589 605 BY SIMILARITY.
 FT DISULFID 607 616 BY SIMILARITY.
 FT DISULFID 619 629 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.

Query Match 6.1%; Score 100; DB 1; Length 3695;
 Best Local Similarity 21.8%; Pred. No. 2.7;
 Matches 57; Conservative 23; Mismatches 100; Indels 82; Gaps 14;

QY 93 ITLAGYCDCAFSDFC-----NNC--NCNNCNLLHDIERFAIKA- 132
 DB 294 ISIGRCVCHGADACADKDPDPFLQCTCQHTCGGTCDCRCPGFNQ--QWKKPATAN 351
 QY 133 -----CLGR-----NPE-----AFQPKIGKQGLGNV---KQHNKGCNCRSGCL 169
 DB 352 SANECQSCNICYGHATDCYYDPEVDRRRASQSLDGTGGGVCIDCQHHTTGVNCR--CL 409
 QY 170 KNYCEYEYAIQICSSSTCKIGKN-----YESPERKTLMSPNV-----MOTGLEG 217
 DB 410 PGFYRSPNHPLOSPHYVCRNCESDFTDGTCDLTGR--CYCRPNFSGERCDCVCAEGFTG 467
 QY 218 SHVLPPTKFSGLPRFSDHRRPSSCISWEVVEATCACLQAQGEAEKHCSCKLAEQMILE 277
 DB 468 FSCYPT-----PSSNDTRQVLPAQIIVNCDCSAAGTQGNACRKD-----P 510
 QY 278 EFGRCLS----QILHTFESKG 295
 DB 511 RVGRCLCKPNFQTHCELCAFG 532

RESULT 14
 ID ITB6_MOUSE STANDARD; PRT; 787 AA.
 AC Q920T9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Integrin beta-6 precursor.
 GN ITB6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RC MEDLINE=20547423; PubMed=11095652;
 RA Arend L.J., Smart A.M., Briggs J.P.;
 RT "Mouse beta(6) integrin sequence, pattern of expression, and role in
 kidney development.";
 RL J. Am. Soc. Nephrol. 11:2297-2305(2000).
 CC -1- FUNCTION: INTEGRIN ALPHA-V/BETA-6 IS A RECEPTOR FOR FIBRONECTIN
 AND CYTOTACTIN. IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGANDS
 (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-6
 CC ASSOCIATES WITH ALPHA-V (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -1- SIMILARITY: Contains 2 VWFA-like domains.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF115376; AAD17212.1; --
 DR HSSP; P05106; 1JY2.
 DR MGD; MGI:96615; Itgb6
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002389; Integrin_B.
 DR InterPro; IPR001169; Integrin_beta_C.
 DR InterPro; IPR003659; Plexin-like.
 DR Pfam; PF00362; Integrin_B.1.
 DR PRINTS; PR01186; INTEGRINB.
 DR ProDom; PD001811; Integrin_B.1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA.2.
 DR PROSITE; PS00222; EGF_1; UNKNOWN_1.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
 KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KW Repeat; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 787 INTEGRIN BETA-6.
 FT DOMAIN 22 706 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 707 729 POTENTIAL.
 FT DOMAIN 730 787 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 131 371 VWFA-LIKE.
 FT DOMAIN 456 619 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 456 501 I.
 FT REPEAT 502 543 II.
 FT REPEAT 544 582 III.
 FT REPEAT 583 619 IV.
 FT DISULFID 23 454 BY SIMILARITY.
 FT DISULFID 31 41 BY SIMILARITY.
 FT DISULFID 34 70 BY SIMILARITY.
 FT DISULFID 44 59 BY SIMILARITY.
 FT DISULFID 197 204 BY SIMILARITY.
 FT DISULFID 252 293 BY SIMILARITY.
 FT DISULFID 394 406 BY SIMILARITY.
 FT DISULFID 426 669 BY SIMILARITY.
 FT DISULFID 452 456 BY SIMILARITY.
 FT DISULFID 467 479 BY SIMILARITY.
 FT DISULFID 476 511 BY SIMILARITY.
 FT DISULFID 481 490 BY SIMILARITY.
 FT DISULFID 492 502 BY SIMILARITY.
 FT DISULFID 517 522 BY SIMILARITY.
 FT DISULFID 519 552 BY SIMILARITY.
 FT DISULFID 524 537 BY SIMILARITY.
 FT DISULFID 539 544 BY SIMILARITY.
 FT DISULFID 558 563 BY SIMILARITY.
 FT DISULFID 560 591 BY SIMILARITY.
 FT DISULFID 565 574 BY SIMILARITY.
 FT DISULFID 576 583 BY SIMILARITY.
 FT DISULFID 597 602 BY SIMILARITY.
 FT DISULFID 599 645 BY SIMILARITY.
 FT DISULFID 604 614 BY SIMILARITY.
 FT DISULFID 617 620 BY SIMILARITY.
 FT DISULFID 624 633 BY SIMILARITY.
 FT DISULFID 630 701 BY SIMILARITY.
 FT DISULFID 649 677 BY SIMILARITY.
 FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 541 541 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 575 575 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 787 AA; 86041 MW; C6438C6F1E6B7FBD CRC64;

Query Match 6.0%; Score 99; DB 1; Length 787;
 Best Local Similarity 27.4%; Pred. No. 0.56;
 Matches 34; Conservative 10; Mismatches 36; Indels 44; Gaps 7;

QY 98 YDCDFASGDFC-----NNCNCNCCNHLHDIERKAIKACLCGRNPEAFQPKIG 146
 Db 573 YCNCTNRDSCSTSEGVLCGRGDCVCGKC-----VC--RNPASGPTCE 615
 QY 147 KGLGNVAPQHNKGCNCRSGCKKNYCEY-----EAQIMSSICKICGCKNYESPCK 201
 Db 616 R-----CPTGDCPCNSKRS-CI-----ECYLSADGQAQECADKCKAIGATISEEDFSKD 664
 QY 202 TLMS 205
 Db 665 TSVS 668

RESULT 15
 LMA2_MOUSE
 ID LMA2_MOUSE STANDARD; PRT; 3106 AA.
 AC Q60675; Q05003; Q64061;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy chain).
 DE LAMA2.
 GN LAMA2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RP [1]
 RC STRAIN=FBV/N; TISSUE=Embryo, and Heart;
 RX MEDLINE=95316259; PubMed=7795883;
 RA Bernier S.M., Utani A., Sugiyama S., Doi T., Polistina C., Yamada Y.
 RT "Cloning and expression of laminin alpha 2 chain (M-chain) in the mouse";
 RL Matrix Biol. 14:447-455(1995).
 RN [2]
 RP SEQUENCE OF 2162-2279 FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RX MEDLINE=93346725; PubMed=8345183;
 RA Chang A.C., Wadsworth S., Coligan J.E.;
 RT "Expression of merosin in the thymus and its interaction with thymocytes";
 RL J. Immunol. 151:1789-1801(1993).
 RN [3]
 RP SEQUENCE OF 64-281 FROM N.A.
 RX MEDLINE=95179178; PubMed=7874173;
 RA Xu H., Wu X.R., Wewer U.M., Engvall E.;
 RT "Murine muscular dystrophy caused by a mutation in the laminin alpha 2 (Lama2) gene";
 RL Nat. Genet. 8:297-302(1994).
 RN [4]
 RP SEQUENCE OF 20-25.
 RX MEDLINE=21818471; PubMed=11829758;
 RA Garbe J.H., Goehring W., Mann K., Timpl R., Sasaki T.;
 RT "Complete sequence, recombinant analysis and binding to laminins and sulphated ligands of the N-terminal domains of laminin alpha3B and alpha5 chains";
 RL Biochem. J. 362:213-221(2002).
 RN [5]
 RP BINDING TO FBLN1, FBLN2, AND NID2.
 RX PubMed=10022829;
 RA Talts J.F., Andac Z., Goehring W., Brancaccio A., Timpl R.;
 RT "Binding of the G domains of laminin alpha1 and alpha2 chains and perlecan to heparin, sulfatides, alpha-dystroglycan and several extracellular matrix proteins";
 RL EMBO J. 18:863-870(1999).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
 RX MEDLINE=20085745; PubMed=10619025;
 RA Hohenester E., Tisi D., Talts J.F., Timpl R.;
 RT "The crystal structure of a laminin G-like module reveals the

RT molecular basis of alpha-dystroglycan binding to laminins, perlecan, and agrin.";
 RL Mol. Cell 4:783-792(1999).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end. The alpha-2 chain is a subunit of laminin-2 (Merosin) and laminin-4 (S-merosin). Interacts with FBLN1, FBLN2 and NID2.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -!- DISEASE: DEFECTS IN LAMA2 ARE A CAUSE OF MURINE MUSCULAR DYSTROPHY (DY2J).
 CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -!- SIMILARITY: Contains 17 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 2 laminin IV domains.
 CC -!- SIMILARITY: Contains 5 laminin G-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12147; AAC52165.1; -;
 DR EMBL; X69869; CAA49502.1; -;
 DR EMBL; S75315; AAB33573.1; -;
 DR PIR; I49077; S53868.
 DR PDB; 1QU0; 03-DEC-99.
 DR PDB; 1DYK; 04-FEB-01.
 DR MGD; MGI:99912; Lama2.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR001886; LamNT.
 DR Pfam; PF00052; laminin_B; 2.
 DR Pfam; PF00053; laminin_EGF; 14.
 DR Pfam; PF00054; laminin_G; 5.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR ProDom; PD003031; Laminin_B; 1.
 DR SMART; SM00180; EGF_Lam; 14.
 DR SMART; SM00281; LamB; 2.
 DR SMART; SM00282; LamG; 5.
 DR SMART; SM00136; LamT; 1.
 DR PROSITE; PS00022; EGF_1; 11.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 5.
 DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 3106 LAMININ ALPHA-2 CHAIN
 FT DOMAIN 20 282 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 283 339 LAMININ EGF-LIKE 1.
 FT DOMAIN 340 409 LAMININ EGF-LIKE 2.
 FT DOMAIN 410 464 LAMININ EGF-LIKE 3.
 FT DOMAIN 465 513 LAMININ EGF-LIKE 4.
 FT DOMAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 720 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT

```
FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.
FT DOMAIN 1056 1101 LAMININ EGF-LIKE 12.
FT DOMAIN 1102 1161 LAMININ EGF-LIKE 13.
FT DOMAIN 1162 1171 LAMININ EGF-LIKE 14.
FT DOMAIN 1172 1375 LAMININ DOMAIN IV 2 (N-TERMINAL).
FT DOMAIN 1376 1415 LAMININ EGF-LIKE 14 (C-TERMINAL).
FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15.
FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.
FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.
FT DOMAIN 1570 2140 DOMAIN II AND I.
FT DOMAIN 2141 2324 LAMININ G-LIKE 1.
FT DOMAIN 2325 2517 LAMININ G-LIKE 2.
FT DOMAIN 2518 2706 LAMININ G-LIKE 3.
FT DOMAIN 2707 2930 LAMININ G-LIKE 4.
FT DOMAIN 2931 3106 LAMININ G-LIKE 5.
FT DOMAIN 3107 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1864 2146 BY SIMILARITY.
FT DISULFID 2147 283 292 BY SIMILARITY.
FT DISULFID 284 303 304 BY SIMILARITY.
FT DISULFID 305 314 315 BY SIMILARITY.
FT DISULFID 316 337 338 BY SIMILARITY.
FT DISULFID 339 349 350 BY SIMILARITY.
FT DISULFID 351 374 375 BY SIMILARITY.
FT DISULFID 376 386 387 BY SIMILARITY.
FT DISULFID 388 407 408 BY SIMILARITY.
FT DISULFID 409 422 423 BY SIMILARITY.
FT DISULFID 424 438 439 BY SIMILARITY.
FT DISULFID 440 449 450 BY SIMILARITY.
FT DISULFID 451 462 463 BY SIMILARITY.
FT DISULFID 464 478 479 BY SIMILARITY.
FT DISULFID 480 482 483 BY SIMILARITY.
FT DISULFID 484 493 494 BY SIMILARITY.
FT DISULFID 495 511 512 BY SIMILARITY.
FT DISULFID 513 522 523 BY SIMILARITY.
FT DISULFID 524 553 554 BY SIMILARITY.
FT DISULFID 555 569 570 BY SIMILARITY.
FT DISULFID 571 772 773 BY SIMILARITY.
FT DISULFID 774 800 801 BY SIMILARITY.
FT DISULFID 802 818 819 BY SIMILARITY.
FT DISULFID 820 828 829 BY SIMILARITY.
FT DISULFID 830 840 841 BY SIMILARITY.
FT DISULFID 842 858 859 BY SIMILARITY.
FT DISULFID 860 875 876 BY SIMILARITY.
FT DISULFID 877 882 883 BY SIMILARITY.
FT DISULFID 884 894 895 BY SIMILARITY.
FT DISULFID 896 911 912 BY SIMILARITY.
FT DISULFID 913 926 927 BY SIMILARITY.
FT DISULFID 928 933 934 BY SIMILARITY.
FT DISULFID 935 944 945 BY SIMILARITY.
FT DISULFID 946 960 961 BY SIMILARITY.
FT DISULFID 962 975 976 BY SIMILARITY.
FT DISULFID 977 981 982 BY SIMILARITY.
FT DISULFID 983 992 993 BY SIMILARITY.
FT DISULFID 994 1007 1008 BY SIMILARITY.
FT DISULFID 1009 1019 1020 BY SIMILARITY.
FT DISULFID 1021 1026 1027 BY SIMILARITY.
FT DISULFID 1028 1037 1038 BY SIMILARITY.
FT DISULFID 1039 1053 1054 BY SIMILARITY.
FT DISULFID 1055 1068 1069 BY SIMILARITY.
FT DISULFID 1070 1075 1076 BY SIMILARITY.
FT DISULFID 1077 1086 1087 BY SIMILARITY.
FT DISULFID 1088 1099 1099 BY SIMILARITY.
FT DISULFID 1100 1425 1425 BY SIMILARITY.
FT DISULFID 1426 1432 1432 BY SIMILARITY.
FT DISULFID 1433 1444 1444 BY SIMILARITY.
FT DISULFID 1445 1447 1447 BY SIMILARITY.
```

```
Matches 70; Conservative 21; Mismatches 82; Indels 145; Gaps 19;
Qy 73 PSVYNGSAPP-----SGSTL---PGPPKITLAGYCDCFAS---GDFC 108
Db 851 PSVPGGSCQCOCNDNLDYSIPGSCDSUSGSLCKPKG---TTGRYCELCADGYFGDAV 906
Qy 109 NNCNCNNC-CN-----NLHHDIERFKAIRACLGRN-----PEAFQPKRIGKQGLGNVVP 155
Db 907 NTKNCPCCRDINGSFSEDCHTRTQCECRPNVQGRHCDCECKPETFGQLGRGCL----- 961
Qy 156 QHNKGCNCRSRGCLKNYCECYEAQIMCS---SICKICGCK---NYEE----- 196
Db 962 ----PCNCNFRGSKSFDCEA-SGQCWCQPGVAGKRCDCRCAHGYNFQEGGCIACDCSHLG 1016
Qy 197 ---SPERKTLMSMPNYMOTGGLEGSHYLPPT----- 224
Db 1017 NNCDPRTQCICPPN---TTGEKCECLPNTWGHISIVTGCKVCNCGSTVGSLASQCNVNTG 1073
Qy 225 -----KFSGLPRFSDHRRPSSCI--SWEV-----VEAT-----CACLIA 256
Db 1074 QCSCHPKFSGM-----KCSECSRGHNNYPLCTLCDGFLPGTDATTCDETRKCSQDO 1126
Qy 257 QGE-----BAEKEHCSC 269
Db 1127 TQCSCKVNVGVGHCDRC 1144
```

Search completed: July 24, 2003, 13:37:35
Job time : 16.101 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:34:27 ; Search time 25.6717 seconds
(without alignments)
1120.082 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVIQLKGGTQMLCIDNSRT.....GRCLSQLHTFEKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	437	26.5	429	2 T23152	hypothetical prote
2	376	22.8	571	2 B84585	hypothetical prote
3	357	21.7	603	2 T08955	hypothetical prote
4	245.5	14.9	658	2 F1410	hypothetical prote
5	117	7.1	497	2 T27827	hypothetical prote
6	112	6.8	1737	2 T00209	MEGF8 protein - hu
7	109.5	6.6	769	1 IJHULM	leukocyte adhesion
8	106.5	6.5	3635	2 T10053	laminin alpha 5 ch
9	105.5	6.4	2180	2 T29764	hypothetical prote
10	104	6.3	775	2 T21436	hypothetical prote
11	103.5	6.3	1786	1 MMFSB1	laminin beta-1 cha
12	102	6.2	1790	1 MMFSB1	laminin beta-1 cha
13	101	6.1	1700	2 S08167	Balbani ring 3 pr
14	101	6.1	1819	2 A71928	cag island protein
15	101	6.1	3011	1 GNWVC3	genome polyprotein
16	100.5	6.1	633	2 T24898	hypothetical prote
17	100	6.1	395	2 S44454	transcription fact
18	100	6.1	520	2 G84510	probable receptor-
19	99.5	6.0	677	2 C42125	trophozoite cystei
20	99	6.0	373	2 T47488	hypothetical prote
21	99	6.0	565	2 T16408	hypothetical prote
22	99	6.0	2195	2 T34264	hypothetical prote
23	98.5	6.0	574	2 B8465	hypothetical prote
24	98.5	6.0	962	2 JC5571	protein B0244.8 [i
25	98.5	6.0	3106	1 S3868	subtilisin-like pr
26	98.5	6.0	5376	2 T42215	laminin alpha-2 ch
27	97.5	5.9	432	2 T37509	zonadhesin - mouse
28	97.5	5.9	969	1 A39490	hypothetical prote
29	97.5	5.9	975	2 JC5570	subtilisin-like pr

30	97.5	5.9	1766	2 A42125	trophozoite cystei
31	97	5.9	382	2 T29339	hypothetical prote
32	97	5.9	798	2 A40526	integrin beta-7 ch
33	97	5.9	3010	1 A45573	genome polyprotein
34	96.5	5.9	290	2 G72858	AcOrf-70 protein -
35	96.5	5.9	1895	2 T15881	hypothetical prote
36	96.5	5.9	3759	2 A35085	trithorax protein
37	96	5.8	798	2 A28193	integrin beta-1 ch
38	96	5.8	3010	1 GNWVC	genome polyprotein
39	95.5	5.8	1801	1 MMRIS	laminin beta-2 cha
40	95	5.8	1927	2 G64585	cag pathogenicity
41	94	5.7	3011	1 GNWVC	genome polyprotein
42	93.5	5.7	772	2 S32659	integrin beta 2 ch
43	93.5	5.7	1599	2 T16210	hypothetical prote
44	93	5.6	1751	1 MMHUMH	laminin alpha-2 ch
45	93	5.6	1810	1 A32230	tenascin precursor

ALIGNMENTS

RESULT 1

T23152
hypothetical protein JC8.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23152
R:Lightning, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19699
A:Accession: T23152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-429 <WIL>
A:Cross-references: EMBL:282274; PIDN:CAB052281.1; GSPDB:GN00022; CESP:JC8.6b
A:Experimental source: clone JC8
C:Genetics:
A:Gene: CESP:JC8.6b
A:Map position: 4
A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Query Match	26.5%	Score 437;	DB 2;	Length 429;
Best Local Similarity	41.3%	Pred. No. 7.2e-28;		
Matches	92;	Conservative 35;	Mismatches 62;	Indels 34; Gaps 8;
Qy	98	YDCDFASDFCNCCNCCNHHDIERFAKACLRNPEAFQPKIG--KGQLGNVKP	155	
Db	182	YDCDFANGFCRDCNCKDCHNNIEYDSQSKAIRSLERNPNAPFKIGIARGGITDIER	241	
Qy	156	QHNKGNCRRSGCLKNYCEYEAQIMCSSICKGCKN-----YEESS-----PERKTL	203	
Db	242	LHQKCHCKKSGCLKNYCEYEAQVPCYDRCKCKGQNTYRTRYNKSGGAVSNTNAL	301	
Qy	204	MSMPNMYQT-----GGL----EGSHYLPPTKFSGLPRFSHRR--PSSCISWVVRAT	250	
Db	302	MSLTNASSTPDSGPGSVWTDHGGDY-EDMLLSHKPKVEMDRPFYMYTDEWVEAA	360	
Qy	251	CACLLAQGEA-----EKEHCSKCLAEQMILEEFGCLSQL	287	
Db	361	TMCWVAQAEALNVEKVTDEKLNMEKLVLRFEGRCLQMI	403	

RESULT 2

B84585
hypothetical protein At2g20110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84585
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-769, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

[illegible]

A:Experimental source: clone 2K287
C:Genetics:
A:Gene: CESP:2K287.1
A:Map position: 5
A:Introns: 20/1; 44/2; 99/3; 339/3; 443/2
Query Match 7.1%; Score 117; DB 2; Length 497;
Best Local Similarity 23.3%; Pred. No. 0.067;
Matches 55; Conservative 26; Mismatches 81; Indels 74; Gaps 13;
QY 99 CDCFASGDFCNN-CNCCNCCNLLHDIERFAIRACLGRRPEAQPQKIGKQOLGNV-KPQ 156
DB 77 CDCURIEGCKSLKACQIC-----KNEA--PK-----KLAKVAKP- 111
QY 157 HNKGCNCR--RSGCLKNYCEYAEQIMCSTICKTG-CKNYEESPERKTLMSMPNMYMTG 213
DB 112 -TSGCQACAGKKQCKREACRTVYGFCASCKGGDCTN-----GASKESVPKHQNC 164
QY 214 GLESHYLPTPKFSGL-----PRFSDRRPSSCSWEVEATCACL----- 254
DB 165 FLEHKH-----ESSGLVTLIGEDVYVRGDFYHESKGEHPVEEQLVAAIYDLISKYTVDL 219
QY 255 -----LAQGEAEAEHCKSLAEQWILEEFCRQLSOLHTFEKSKGLKM 298
DB 220 HEIQIFVSKSPCFHDCPEPCVEVDECKSNKACAKLLGLLSKV-RKEIKKVDVKM 274
RESULT 6
T00209
MEGF8 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00209
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: 214126; MUID:98360089; PMID:9693030
A:Accession: T00209
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1737 <NAK>
A:Cross-references: EMBL:AB011541; NID:g3449307; PIDN:BA32469.1; PID:g3449308
A:Experimental source: brain; clone HGI392
C:Genetics:
A:Gene: MEGF8
A:Map position: 19q12

Query Match 6.8%; Score 112; DB 2; Length 1737;
Best Local Similarity 20.2%; Pred. No. 0.58;
Matches 51; Conservative 33; Mismatches 73; Indels 96; Gaps 15;
QY 86 TLPGPKITLAGYCDCAFSGDFCNN---CN-CNNCC---NNLH-----H 122
DB 763 TLPDDP-----CRLSSPACNQSGACTWCHGACLSGDAQHRLCGGSPCSPWRSP 815
QY 123 DIERFKAIKACLGRRNPEAFQPKIGKQOLGNVKKQHNKGCNRRSGCL-----KNYCE 174
DB 816 ECRRLRTCSCLARHPTLQP--GDGEAST--PRCKWCTNCPGACIGRNGSCTSEND 871
QY 175 CYEAQIMCSSTC-----KCIGKNYEEESPERKTLMSM-PNY----- 209
DB 872 INQREVFWAGNCSEAACGAADCEQCTREGKCMWTRQFRRTGETRILSVQPTDYDTCFSH 931
QY 210 ---MOTGGLGSHVLP-PTKFSGLPFRSHDRRPSSCISWEVEATCACLAAQGEAEKE 264
DB 932 SLLNVSPMVESPPLPCTPCHLLP-----NCTS-----CLDSKGDAGGNQ 973
QY 265 HC-----SKCIA 271
DB 974 HCVWSSSLQQLS 986

RESULT 7

IJHULM
leukocyte adhesion protein beta chain (CD18) precursor - human
N:Alternate names: integrin beta-2; LFA-1 beta chain; Mac-1 beta chain; p150,95 beta
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 23-Jul-1999
C:Accession: A25967; A91084; S19324; I70090; I70091; A29265; A94497
R:Kishimoto, T.K.; O'Connor, K.; Lee, A.; Roberts, T.M.; Springer, T.A.
Cell 48, 681-690, 1987
A:Title: Cloning of the beta subunit of the leukocyte adhesion proteins: homology to
A:Reference number: A25967; MUID:87131080; PMID:3028646
A:Accession: A25967
A:Molecule type: mRNA
A:Residues: 1-769 <KTS>
A:Cross-references: GB:M15395; NID:g186933; PIDN:AAA59490.i; PID:g307113
A:Note: source of LFA-1 was the SKW3 T-cell line; source of Mac-1 was pooled leukocyte
B:Law, S.K.A.; Gagnon, J.; Hildreth, J.E.; Wells, C.E.; Willis, A.C.; Wong, A.J.
EMBO J. 6, 915-919, 1987
A:Title: The primary structure of the B-subunit of the cell surface adhesion glycopro
A:Reference number: A91084; MUID:87246525; PMID:2954816
A:Accession: A91084
A:Molecule type: mRNA
A:Residues: 9-198, 'P', 200-769 <LA2>
A:Cross-references: GB:Y00057; NID:g30228; PIDN:CAA68266.1; PID:g762939
R:Weltman, J.B.; Wells, C.E.; Wright, A.H.; Clark, P.A.; Law, S.K.A.
FEBS Lett. 294, 97-103, 1991
A:Title: The gene organisation of the human beta2 integrin subunit (CD18).
A:Reference number: S19324; MUID:92077153; PMID:1683838
A:Accession: S19324
A:Molecule type: DNA
A:Residues: 1-3;16-23;46-53;106-113;163-170;244-251;296-303;328-335;358-365;405-412;4
R:Nelson, C.; Rabb, H.; Arnaout, M.A.
J. Biol. Chem. 267, 3351-3357, 1992
A:Title: Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing a
A:Reference number: I55376; MUID:92147694; PMID:1346613
A:Accession: I70090
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 347-350, 'S', 352-355 <NEU>
A:Cross-references: GB:S81234; NID:g245294; PIDN:AAB21404.1; PID:g245295
A:Note: mutant form
A:Accession: I70091
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 246-247, 'PSSO', 248-249 <RES>
A:Cross-references: GB:S81252; NID:g245289; PIDN:AAB21402.1; PID:g245290
A:Note: mutant splice form
C:Comment: The leukocyte adhesion proteins are noncovalently linked heterodimers of d
and a deficiency of them is attributed to a genetic defect in the expression or stru
C:Comment: The cysteine residues are involved in intrachain disulfide bonds.
C:Genetics:
A:Gene: GDB:ITGB2
A:Cross-references: GDB:120574; OMIM:600065
A:Map position: 21q22.3-21q22.3
C:Superfamily: integrin beta chain; laminin-type EGF-like homology
C:Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; leuk
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-769/Product: leukocyte adhesion protein beta chain #status predicted <MAT>
F:23-670/Domain: extracellular #status predicted <EXT>
F:445-631/Region: cysteine-rich
F:459-540/Region: duplication
F:541-627/Region: duplication
F:701-723/Domain: transmembrane #status predicted <MEM>
F:724-769/Domain: intracellular #status predicted <CYT>
F:23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic
F:50,116,212,254,501/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 6.6%; Score 109.5; DB 1; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.42;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVICQLKGTQMLCIDNSRTRELKALHLVPQYQDQNNYLOSVPKPMALVGRFLPASTK 60
DB 436 IYTVQVLPQCECRCDQSRDRSL--CH-----GKGFLECGICRCDTGYIGKNCEQIQ 486

A;Residues: 1-2180 <DU2>
A;Cross-references: EMBL:AF003133; PIDN:AA554138.1; GSPDB:GN00019; CESP:T21E3.3
A;Experimental source: strain Bristol N2; clone T21E3
C;Genetics:
A;Gene: CESP:T21E3.3
A;Map position: 1
A;Introns: 400/2; 1608/3; 1644/1; 1674/3; 1686/2; 1731/2; 2073/1; 2098/2; 2146/3
C;Superfamily: LDL receptor ligand-binding repeat homology
F;15-49/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;51-85/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;90-132/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;149-185/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;190-225/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F;229-272/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F;285-316/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F;323-359/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F;369-405/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F;816-856/Domain: LDL receptor ligand-binding repeat homology <LDL10>
F;861-903/Domain: LDL receptor ligand-binding repeat homology <LDL11>
F;908-943/Domain: LDL receptor ligand-binding repeat homology <LDL12>
F;948-987/Domain: LDL receptor ligand-binding repeat homology <LDL13>
F;993-1027/Domain: LDL receptor ligand-binding repeat homology <LDL14>
F;1031-1063/Domain: LDL receptor ligand-binding repeat homology <LDL15>
F;1126-1165/Domain: LDL receptor ligand-binding repeat homology <LDL16>
F;1172-1209/Domain: LDL receptor ligand-binding repeat homology <LDL17>
F;1213-1246/Domain: LDL receptor ligand-binding repeat homology <LDL18>

Query Match 6.4%; Score 105.5; DB 2; Length 2180;
Best Local Similarity 21.7%; Pred. No. 2.4;
Matches 67; Conservative 22; Mismatches 115; Indels 105; Gaps 13;

QY 25 ALHLVPQYQDQNNYLSQSDVPKPMALTALVGRFLPASTKLNLITQQL-----TTLNPAAQTFVKDFCECDKGWTGPH 1743
DB 1697 ARHVQFKNSDQQMLSVTAQ-----TTLNPAAQTFVKDFCECDKGWTGPH 1743

QY 69 --EGALPSVYVNGSAFSPSGSTLPQPPKITYAGYCDCFASGDFCNCCNCCN----- 118
DB 1744 CRHKADAKVCYGHCF--SGGACDGEGLPLNLRCSGDLGTNRCONCVGHECLNGGFCSYAN 1802

QY 119 ---NLHDDT-----ERFKAIKACLGRLNPEAPQPKIGKGLGNVKNPKHNGKNCRR 165
DB 1803 SNRSLPHCIKPSGFTGDHCEEYLCKDAC-----PFGSKTYDITRPMDDPITCSCEQ 1853

QY 166 SGCLKNYCEYEAQIMCSSTCKIGCKNYEESPERKTLMSMPNYMTGG--LEGSHYLP 223
DB 1854 NAAAHN-----TDCSPICQ-----KQPNWCHNGGRCLDTPGY--P 1886

QY 224 TKFSGLPFRSHDR-----RPSSCISWVVEATCACLLA--QGEBAEKE-HGSKC 269
DB 1887 GKCKCLPRAFGPCDVPVQCDDYCTNNSKCTITNGTHFECDCKPGFKGLRCEQETKSEC 1946

QY 270 LAEQMLTEE 278
DB 1947 SNEAKCIKK 1955

RESULT 10
T21436
hypothetical protein R06A4.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T21436; T23953
R;Barlow, K.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19421
A;Accession: T21436
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-775 <WIL>
A;Cross-references: EMBL:281515; PIDN:CA804199.1; GSPDB:GN00020; CESP:R06A4.7
R;Barlow, K.
submitted to the EMBL Data Library, November 1996

A:Reference number: A02870; MUID:85051302; PMID:6209134
A:Accession: A02871
A:Molecule type: mRNA
A:Residues: 1292-1530, 'MEMP', 1535-1691, 'C', 1693-1748, 'N', 1750-1786 <BAR>
A:Cross-references: EMBL:X05212; NID:g52861; PIDN:CAA28839.1; PID:g809042
R:Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaeumer, I.; Hartl, L.
Eur. J. Biochem. 177, 35-45, 1988
A:Title: Structural study of long arm fragments of laminin. Evidence for repetitive C
A:Reference number: S01790; MUID:89030693; PMID:3181157
A:Accession: S02036
A:Molecule type: protein
A:Residues: 1561-1587 <DEU>
R:Faulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.
EMBO J. 4, 309-316, 1985
A:Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.
A:Reference number: S13543; MUID:85257455; PMID:3848400
A:Accession: S13543
A:Molecule type: protein
A:Residues: 1700-1748, 'N', 1750-1759 <PAU>
C:Genetics:
A:Gene: Lamb-1
A:Map position: 12
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
C:Function:
A:Description: Interact with cells and with other basement membrane proteins to promote
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1786/Product: laminin beta-1 chain #status predicted <MAT>
F:22-270/Domain: VI <DOM6>
F:271-540/Domain: V <DOM5>
F:271-332/Domain: laminin-type EGF-like homology <LE01>
F:335-395/Domain: laminin-type EGF-like homology <LE02>
F:398-455/Domain: laminin-type EGF-like homology <LE03>
F:458-507/Domain: laminin-type EGF-like homology <LE04>
F:510-540/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:541-772/Domain: IV <DOM4>
F:773-1182/Domain: III <DOM3>
F:773-818/Domain: laminin-type EGF-like homology <LE06>
F:821-864/Domain: laminin-type EGF-like homology <LE07>
F:867-914/Domain: laminin-type EGF-like homology <LE08>
F:917-973/Domain: laminin-type EGF-like homology <LE09>
F:976-1025/Domain: laminin-type EGF-like homology <LE10>
F:1028-1081/Domain: laminin-type EGF-like homology <LE11>
F:1084-1129/Domain: laminin-type EGF-like homology <LE12>
F:1132-1176/Domain: laminin-type EGF-like homology <LE13>
F:1183-1397/Domain: II <DOM2>
F:1183-1397/Region: heptad repeats
F:1398-1430/Domain: alpha <ALP>
F:1431-1786/Region: heptad repeats
F:1431-1786/Domain: I <DOM1>
F:22/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:30-35/Disulfide bonds: #status predicted
F:120-356,519,677,1041,1195,1279,1336,1343,1487,1533,1542,1643/Binding site: carbohyd
F:1179,1182,1785/Disulfide bonds: Interchain #status predicted

Query Match 6.3%; Score 103.5; DB 1; Length 1786;
Best Local Similarity 21.4%; Pred. No. 2.9;
Matches 54; Conservative 29; Mismatches 74; Indels 95; Gaps 15;

QY 90 PPKITLAGYCDCFASGDFCNCC-----NCNNCCNNLHHDIERFAIRACLGR 136
Db PVTQLQACVDFCYIGSRCDACSGFFGNSDFGSGQPC-QCHHNIDT-----T 985
QY 137 NPEAFOPKTK-----GQLGNVKPQHNKGCNRRSCLKNY-----C 173
Db DPEACDKDTGRCLLKCLYHTEGDHQCLQCYGYGQDALRQCRKCVCNVLTGVREHCNGSDC 1045
QY 174 ECYEAQIMCSSICKICGCKNYEESPERKTLMSWPNYMQ---TG----CLEGSHYLPPT- 224
Db 1046 HCDKATGQCSCPNVTVG-----QNCDR----CAPNTWQLASGTGCGPCNCNAHSGPSC 1096
QY 225 -KFSG----LPFRSHDRPSSC--ISW-----EVVEATCACLAAO 257

Db 1097 NEFTGQCQCPGFC-GRTCEQCLFWGDPDVEGRACDCDPRIETPQCQSTGQCVCVE 1155
QY 258 GEEAEKEHCKSC 269
Db 1156 GVEGPR--CDKC 1165

RESULT 12
MMF2B1
laminin beta-1 chain precursor - fruit fly (Drosophila melanogaster)
N:Alternate names: laminin chain B1
C:Species: Drosophila melanogaster
C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C:Accession: A28783; S14462; B28783
R:Montell, D.J.; Goodman, C.S.
Cell 53, 463-473, 1988

A:Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals do
A:Reference number: A28783; MUID:88210471; PMID:3365769
A:Accession: A28783
A:Molecule type: mRNA
A:Residues: 1-1790 <MON1>
A:Cross-references: EMBL:M19525
R:Montell, D.J.; Goodman, C.S.
submitted to the EMBL Data Library, June 1988
A:Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reve
A:Reference number: S14462
A:Accession: S14462
A:Molecule type: mRNA
A:Residues: 1-667, 'L', 669-725, 'VT', 728-947, 950-1790 <MON2>
A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
C:Genetics:
A:Gene: lamB1
A:Cross-references: FlyBase:FBgn0002527
A:Map position: 2L 28D
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C:Function:
A:Description: Interact with cells and with other basement membrane proteins to promote
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
F:27-286/Domain: VI <DOM6>
F:289-561/Domain: V <DOM5>
F:290-354/Domain: laminin-type EGF-like homology <LE01>
F:357-417/Domain: laminin-type EGF-like homology <LE02>
F:420-477/Domain: laminin-type EGF-like homology <LE03>
F:480-528/Domain: laminin-type EGF-like homology <LE04>
F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:562-789/Domain: IV <DOM4>
F:643-645/Region: cell attachment (R-G-D) motif
F:790-1189/Domain: III <DOM3>
F:791-836/Domain: laminin-type EGF-like homology <LE06>
F:839-882/Domain: laminin-type EGF-like homology <LE07>
F:885-932/Domain: laminin-type EGF-like homology <LE08>
F:935-990/Domain: laminin-type EGF-like homology <LE09>
F:968-972/Region: cell adhesion #status predicted
F:993-1042/Domain: laminin-type EGF-like homology <LE10>
F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
F:1190-1407/Domain: II <DOM2>
F:1408-1434/Domain: alpha <ALP>
F:1435-1790/Domain: I <DOM1>
F:51-56/Disulfide bonds: #status predicted
F:140_203_234_489_593_1053_1248_1303_1332_1343_1475_1495_1517_1583_1646_1705/Binding sit
F:1191_1194_1788/Disulfide bonds: Interchain #status predicted

Query Match 6.2%; Score 102; DB 1; Length 1790;
Best Local Similarity 21.4%; Pred. No. 3.8;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

QY 97 GYCD-----CF--ASGDFCNCC-----NCNCC-----CNLLHHDIERFKA 129

Db 1005 GNCDROTGACLKLYOTTDGHCCLCKDFFGDALQONCQCCDFLGTNTNTIAHCDRFTG 1064
QY 130 IKACILGRNPEAFQPKIGKGQGNVKKQHNK-----GCNRRSGCLKNKCCEYEAQIM 181
Db 1065 QCPCL-----PNVOGVRCDOCAENHWKIASGEGCESCNCDDPIGALHQCNSYTGCCQ 1116
QY 182 CS---SICKCIGCK-NYEESPERKTLMSMPNMYMTGLEGSHYLPPTKFS--GLPRFSD 235
Db 1117 CKPFGGRACNCOAHYGNPNEK-----CQCECOFGAADFQCD 1157
QY 236 RPPSCISWEVVEATCACLAAQGEAEKHCSCKLADQMILEBFGRLS 284
Db 1158 RETGNCVCHGIGGYKNECARGYIGQPHCSPC-----GECFN 1196

RESULT 13
S08167
Balbiani ring 3 protein - midge (Chironomus tentans)
C:Species: Chironomus tentans
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 21-Jul-2000
C:Accession: S08167
R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.
J. Mol. Biol. 211, 331-349, 1990
A:Title: The balbiani ring 3 gene in Chironomus tentans has a diverged repetitive str
A:Reference number: S08167; MUID:90172404; PMID:1689777
A:Accession: S08167
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1700 <PAU>
A:Cross-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058
C:Genetics:
A:Gene: BR3
A:Map position: 4
C:Superfamily: unassigned Balbiani ring proteins

Query Match 6.1%; Score 101; DB 2; Length 1700;
Best Local Similarity 22.6%; Pred. No. 4.4;
Matches 51; Conservative 20; Mismatches 87; Indels 68; Gaps 15;

QY 99 CDCFASGD-----FC-NNCNC-----NCCNLLHHDIERFKAICAGLRNPEAF 141
Db 198 CSCECKGDKGCKQSKIWCKNKNCRICPTAEPAGCSAPLKWDDD--KSCACAPAKMEKK 255
QY 142 QPKIGRGQLGNVRKQHNKGCNRRSGCLKN-----YCECYE-----AQIMCSICK 187
Db 256 EKCVEGSKIWN---PNTCEGCAQLNCPDNKANKKETCQCEKEVKKCGQGVCKDCSCS 312
QY 188 CIGCKNYEESPERKTLMSMPNMYMTGLEGSHYLP---PTKFSGLPR----- 231
Db 313 CV-CPGGDKD---KTCTAPQVY---DGVACSCSPVNMOKPADGCPRPQKWDKEECREC 365
QY 232 -FSHRRPSSCISWEVVEATCAC-----LLAOGEEAEKHCSC-KCL 270
Db 366 PVKEDCKNGV--WDETICQICPRDAPVCTAGKERGESCECKCI 409

RESULT 14
A71928
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: A71928
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:37:08 ; Search time 19.6313 Seconds
(without alignments)
1808.803 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQLHTEFKSLKIME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published_Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	446	27.1	147	11	US-09-764-891-3988
2	437	26.5	438	9	US-09-220-091-13
3	278	16.9	53	15	US-10-106-698-6932
4	112.5	6.8	1192	15	US-10-189-971-18
5	112.5	6.8	1207	15	US-10-189-971-20
6	112.5	6.8	1477	15	US-10-189-971-8
7	112.5	6.8	1535	15	US-10-189-971-14
8	112.5	6.8	1593	15	US-10-189-971-4
9	112.5	6.8	4588	15	US-10-137-129A-3
10	109.5	6.6	769	14	US-10-072-841-31
11	109	6.6	1653	15	US-10-133-172-20
12	106.5	6.5	3635	10	US-09-845-583-2
13	106.5	6.5	3635	15	US-10-037-182-4
14	105.5	6.4	695	9	US-09-864-761-42967
15	105	6.4	1057	15	US-10-189-971-6

16	105	6.4	1251	15	US-10-189-971-16	Sequence 16, Appl
17	105	6.4	1342	15	US-10-189-971-24	Sequence 24, Appl
18	105	6.4	1512	15	US-10-189-971-10	Sequence 10, Appl
19	105	6.4	1570	15	US-10-189-971-12	Sequence 12, Appl
20	105	6.4	1628	15	US-10-189-971-2	Sequence 2, Appl
21	104	6.3	773	15	US-10-231-778-223	Sequence 223, Appl
22	103.5	6.3	1725	15	US-10-037-182-12	Sequence 12, Appl
23	103.5	6.3	1786	15	US-09-938-275-7	Sequence 7, Appl
24	103.5	6.3	1786	15	US-10-037-182-10	Sequence 10, Appl
25	102.5	6.2	4123	15	US-10-213-509-5	Sequence 5, Appl
26	101.5	6.2	686	10	US-09-881-654-2	Sequence 2, Appl
27	101.5	6.2	686	10	US-09-881-239-3	Sequence 3, Appl
28	101.5	6.2	728	10	US-09-881-239-1	Sequence 1, Appl
29	101	6.1	465	16	US-10-232-643-2	Sequence 2, Appl
30	101	6.1	631	16	US-10-232-643-1	Sequence 1, Appl
31	101	6.1	1099	10	US-09-881-654-4	Sequence 4, Appl
32	101	6.1	2894	10	US-09-941-611-23	Sequence 23, Appl
33	101	6.1	2894	15	US-10-044-995-23	Sequence 23, Appl
34	101	6.1	3011	9	US-09-916-359-2	Sequence 2, Appl
35	100.5	6.1	760	15	US-10-231-778-224	Sequence 224, App
36	100.5	6.1	1037	15	US-10-152-724A-3	Sequence 3, Appl
37	100	6.1	1416	15	US-10-133-172-4	Sequence 4, Appl
38	100	6.1	2743	15	US-10-037-182-36	Sequence 36, Appl
39	100	6.1	3695	15	US-10-037-182-2	Sequence 2, Appl
40	99.5	6.0	1551	9	US-09-970-318-4	Sequence 4, Appl
41	98.5	6.0	2749	15	US-10-123-155-385	Sequence 385, App
42	98.5	6.0	2749	16	US-10-146-731-385	Sequence 385, App
43	97.5	5.9	632	10	US-09-929-955-29	Sequence 29, Appl
44	97.5	5.9	686	10	US-09-929-955-17	Sequence 17, Appl
45	97.5	5.9	686	10	US-09-929-955-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-09-764-891-3988
; Sequence 3988, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3988
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3988

Query Match	27.18;	Score 446;	DB 11;	Length 147;
Best Local Similarity	98.98;	Pred. No. 2.7e-36;		
Matches	87;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
Qy	1	MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDNVYQSDVDPKPMYALVGRFLPASTK 60		
Db	60	MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDNVYQSDVDPKPMYALVGRFLPASTK 119		
Qy	61	LNLTITQLEGALPSVWNGSAPPSGSLP 88		
Db	120	LNLTITQLEGALPSVWNGSAPPSGSTXP 147		
RESULT 2				
US-09-220-091-13				

```
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Query Match          26.5%; Score 437; DB 9; Length 438;
Best Local Similarity 41.3%; Pred. No. 9e-35;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;

QY 98 YDCDFASGDFCNCCNCCNHLHDIERFAKACLAGRNPEAFQPKIG--KGQLGNVKP 155
DB 191 YDCDFANGFCRDCNCKCHNIEYDSQSKAIQROSLERNPNAFKIGIARGGITDIER 250
QY 156 QHNKGCNRRSCLNICYECYBAQIMCSSICKICGCKN-----YEES-----PERKTL 203
DB 251 LHQKGCHCKKSGCLNICYECYBAQVPCIDRCKCGQNTETRYMTRYKNSGGAVSNTNAL 310
QY 204 MSMPNYMOT-----GGL-----EGSHYLPPTKFSCLPFSHDRR--PSSCLISWEVVEAT 250
DB 311 MSLTNASSTATPDSPGSGVVTDEHGDDY--EDMLLSHKPKVMDPRRFPWYMTDEVVEAA 369
QY 251 CACLLAQGBEA-----EKEHCKSKLAEQMILEEFGRCLSQIL 287
DB 370 TWCWVAQAEALNYEKVQTEDEKLNMEKLVLRREFGRCLQEMI 412

RESULT 3
US-10-106-698-6932
; Sequence 6932, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6932

Query Match          16.9%; Score 278; DB 15; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.8e-20;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 VEATCACLAAQGEAEKEHCKSKLAEQMILEEFGRCLSQILHTEFKSKGLKME 299
|||||

; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Query Match          26.5%; Score 437; DB 9; Length 438;
Best Local Similarity 41.3%; Pred. No. 9e-35;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;

QY 98 YDCDFASGDFCNCCNCCNHLHDIERFAKACLAGRNPEAFQPKIG--KGQLGNVKP 155
DB 191 YDCDFANGFCRDCNCKCHNIEYDSQSKAIQROSLERNPNAFKIGIARGGITDIER 250
QY 156 QHNKGCNRRSCLNICYECYBAQIMCSSICKICGCKN-----YEES-----PERKTL 203
DB 251 LHQKGCHCKKSGCLNICYECYBAQVPCIDRCKCGQNTETRYMTRYKNSGGAVSNTNAL 310
QY 204 MSMPNYMOT-----GGL-----EGSHYLPPTKFSCLPFSHDRR--PSSCLISWEVVEAT 250
DB 311 MSLTNASSTATPDSPGSGVVTDEHGDDY--EDMLLSHKPKVMDPRRFPWYMTDEVVEAA 369
QY 251 CACLLAQGBEA-----EKEHCKSKLAEQMILEEFGRCLSQIL 287
DB 370 TWCWVAQAEALNYEKVQTEDEKLNMEKLVLRREFGRCLQEMI 412

RESULT 3
US-10-106-698-6932
; Sequence 6932, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6932

Query Match          16.9%; Score 278; DB 15; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.8e-20;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 VEATCACLAAQGEAEKEHCKSKLAEQMILEEFGRCLSQILHTEFKSKGLKME 299
|||||

Db 1 VEATCACLAAQGEAEKEHCKSKLAEQMILEEFGRCLSQILHTEFKSKGLKME 53

RESULT 4
US-10-189-971-18
; Sequence 18, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1e1 Human Kielin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 1192
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-18

Query Match          6.8%; Score 112.5; DB 15; Length 1192;
Best Local Similarity 20.0%; Pred. No. 0.039;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;

QY 78 GSAPFSGSTLPQP-----PKITLAGYCDCFASGDFCNCCN-----112
DB 326 GREFASGERFSPSTAACHLCLCWESGVSCEPRACAPALCPFPARGDCDCDCEVLGSH 385
QY 113 -----CNCCNHLHDIERFAKACLAG-----RNPEAFQPKIGKQ 149
DB 386 YLSNQEFDPDPKPCNL-----TCLGGFVTCGRRPCEPPGCSHPLIPSGH 430
QY 150 L-----GNVKP-----QHNKGCNRRSCLNICYECYEA 178
DB 431 CQPTCQGRYHGVTTASGETLPDLPDPTCSLCTCGREHQDEFEFGPAGSCWCKCQAG 490
QY 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPNYMOTGGLEGSH 219
DB 491 QVSCVRLQCPPLPCKLQVTERGSCCPRCRGCLAHGEEHP-----EGSR 533
QY 220 YLPPTKFSGLPRFSDRRPSSCISWEVVEATCA---CL---LAQGEAEKEHCKSK 269
DB 534 WVPP-----DSACSCVCHEGV-VTCARIQCISSCAQPRQGHDPHDCPCQC 576

RESULT 5
US-10-189-971-20
; Sequence 20, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1e1 Human Kielin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
```

LENGTH: 1207
TYPE: PRT
ORGANISM: homo sapiens
US-10-189-971-20

Query Match 6.8%; Score 112.5; DB 15; Length 1207;
Best Local Similarity 20.0%; Pred. No. 0.04;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;
Qy 78 GSAPFSGSTLPGP-----PKITLAGYCDGCFASGDFCNCCN-----112
Db 341 GKEFASGERFSPPTAAACHLCWEGSVSCEPKACAPALCPPARGDCPCDCGCEYLGES 400
Qy 113 -----CNCCNNLHDIERFKAIKACLG-----RNPEAFQPKIGKQ 149
Db 401 YLSNQEFPPDPCNLC-----TCLGFEVTCGRPPCGCSHPLIPSGH 445
Qy 150 L-----GNVKP-----OHKGCNCRSGCLKNKYCEYEA 178
Db 446 CCPTCGCRYHGVTTASGETLPDPTCSLCTCGREHODGEFEFGPAGSCWCRCQAG 505
Qy 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPNMYQTGLGESH 219
Db 506 QVSCVRLQCPPLCKLQVTERGSCCPRRCGLAHGEHP-----EGSR 548
Qy 220 YLPPTKFSGLPRFSDRRPSSCISWEVVEATCA---CL--LAQGEAEKEHCKSKC 269
Db 549 WVPP-----DSACSSCVCHGV-VTCARIQCISSCAQPRQGHDDCCPQC 591

RESULT 6

US-10-189-971-8
Sequence 8, Application US/10189971
Publication No. US20030028907A1
GENERAL INFORMATION:

APPLICANT: Walke, D. Wade
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
FILE REFERENCE: LEX-0360-USA
CURRENT APPLICATION NUMBER: US/10/189,971
CURRENT FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/302,949
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: US 60/315,634
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8

LENGTH: 1477

TYPE: PRT

ORGANISM: homo sapiens
US-10-189-971-8

Query Match 6.8%; Score 112.5; DB 15; Length 1477;
Best Local Similarity 20.0%; Pred. No. 0.052;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;
Qy 78 GSAPFSGSTLPGP-----PKITLAGYCDGCFASGDFCNCCN-----112
Db 611 GKEFASGERFSPPTAAACHLCWEGSVSCEPKACAPALCPPARGDCPCDCGCEYLGES 670
Qy 113 -----CNCCNNLHDIERFKAIKACLG-----RNPEAFQPKIGKQ 149
Db 671 YLSNQEFPPDPCNLC-----TCLGFEVTCGRPPCGCSHPLIPSGH 715
Qy 150 L-----GNVKP-----OHKGCNCRSGCLKNKYCEYEA 178
Db 716 CCPTCGCRYHGVTTASGETLPDPTCSLCTCGREHODGEFEFGPAGSCWCRCQAG 775
Qy 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPNMYQTGLGESH 219

Db 776 QVSCVRLQCPPLCKLQVTERGSCCPRRCGLAHGEHP-----EGSR 818
Qy 220 YLPPTKFSGLPRFSDRRPSSCISWEVVEATCA---CL--LAQGEAEKEHCKSKC 269
Db 819 WVPP-----DSACSSCVCHGV-VTCARIQCISSCAQPRQGHDDCCPQC 861

RESULT 7

US-10-189-971-14
Sequence 14, Application US/10189971
Publication No. US20030028907A1
GENERAL INFORMATION:

APPLICANT: Walke, D. Wade
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
FILE REFERENCE: LEX-0360-USA
CURRENT APPLICATION NUMBER: US/10/189,971
CURRENT FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/302,949
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: US 60/315,634
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 1535
TYPE: PRT
ORGANISM: homo sapiens
US-10-189-971-14

Query Match 6.8%; Score 112.5; DB 15; Length 1535;
Best Local Similarity 20.0%; Pred. No. 0.055;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;
Qy 78 GSAPFSGSTLPGP-----PKITLAGYCDGCFASGDFCNCCN-----112
Db 669 GKEFASGERFSPPTAAACHLCWEGSVSCEPKACAPALCPPARGDCPCDCGCEYLGES 728
Qy 113 -----CNCCNNLHDIERFKAIKACLG-----TCLGFEVTCGRPPCGCSHPLIPSGH 773
Db 729 YLSNQEFPPDPCNLC-----GNVKP-----OHKGCNCRSGCLKNKYCEYEA 178
Qy 150 L-----GNVKP-----OHKGCNCRSGCLKNKYCEYEA 178
Db 774 CCPTCGCRYHGVTTASGETLPDPTCSLCTCGREHODGEFEFGPAGSCWCRCQAG 833
Qy 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPNMYQTGLGESH 219
Db 834 QVSCVRLQCPPLCKLQVTERGSCCPRRCGLAHGEHP-----EGSR 876

Qy 220 YLPPTKFSGLPRFSDRRPSSCISWEVVEATCA---CL--LAQGEAEKEHCKSKC 269
Db 877 WVPP-----DSACSSCVCHGV-VTCARIQCISSCAQPRQGHDDCCPQC 919

RESULT 8

US-10-189-971-4
Sequence 4, Application US/10189971
Publication No. US20030028907A1
GENERAL INFORMATION:

APPLICANT: Walke, D. Wade
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
FILE REFERENCE: LEX-0360-USA
CURRENT APPLICATION NUMBER: US/10/189,971
CURRENT FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/302,949
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: US 60/315,634

```
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-4

Query Match          6.8%; Score 112.5; DB 15; Length 1593;
Best Local Similarity 20.08; Pred. No. 0.058;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;

Qy 78 GSAPPSCSTLPGP-----PKITLAGYCDCFASGDFCNMNCN-----112
Db 727 GKEFASGERPSPTAACHLCLWEGSVCEPKACAPALCPFPARGDCDCDCEYLGES 786
Qy 113 -----CNCCNNLHHDIERFAKACLG-----RNPAFQPKIGKQ 149
Db 787 YLSNOEFPDPREPCNLG-----TCLGFGVTCGRPRCPPEPGCSHPLIPSGH 831
Qy 150 L-----GNVKP-----OHNKGCNCRSGCLKNYCEYEA 178
Db 832 CCPTQCGRYGVTTASGETLPDPLDPTCSLCTCQGREHODGEEFEGPAGSCWCRCQAG 891
Qy 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPTMOTGGLEGSH 219
Db 892 QVSCVRLQCPPLCKLQVTERGSCCPRCGLAHGEHP-----EGSR 934
Qy 220 YLPPTKFSGLPRFSDHRRSPSCISWEVEATCA---CL--LAQEEAEKHEHCKC 269
Db 935 WVPP-----DSACSSVCHEGV-VTCARIQCISSCAQPRQGHDCCPQC 977

RESULT 9
US-10-137-129A-3
; Sequence 3, Application US/10137129A
; Publication No. US20030124108A1
; GENERAL INFORMATION:
; APPLICANT: Frohman, Michael
; APPLICANT: Morris, Andrew
; APPLICANT: Engebrecht, Joanne
; FILE REFERENCE: Onyx2004-DIV2
; CURRENT FILING DATE: 2003-01-08
; CURRENT APPLICATION NUMBER: US/10/137.129A
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4588
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-137-129A-3

Query Match          6.8%; Score 112.5; DB 15; Length 4588;
Best Local Similarity 21.28; Pred. No. 0.24;
Matches 52; Conservative 14; Mismatches 94; Indels 85; Gaps 10;

Qy 49 ALVGRFLPASTKUNLITQLEGALPSVVGNSAFPSGTLPG-----PPKITLAGY 98
Db 2179 AAAGGALPIQKSIDVDVSKLGA---TAGGAA--AGCCAAGAAAGTTCTCCAAATTTAGT 2233
Qy 99 CDCFASGDFCNCCNCCNNLHHDIERFAKACLRNPEAFQPKIGKQLGNVKPQHN 158
Db 2234 CTCTACAGGAGCTGICGPKRF-SKFSLYKOLCACAG-----2269
Qy 159 KGCNCR--SGCLKNYCEYEAQIMCSSICKICGKNYEESEPERKTLMSMPTMOTGGLE 216
Db 2270 -GCACCACTGCAGCAGCAGATAGCATCAGCAT-----TGACA 2310
Qy 217 GSHYLPPTKFSGLPRFSDHRRSPSCI-----SWEVEATCACLIAQEEAEKE 264

; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-4

Query Match          6.8%; Score 112.5; DB 15; Length 1593;
Best Local Similarity 20.08; Pred. No. 0.058;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;

Qy 78 GSAPPSCSTLPGP-----PKITLAGYCDCFASGDFCNMNCN-----112
Db 727 GKEFASGERPSPTAACHLCLWEGSVCEPKACAPALCPFPARGDCDCDCEYLGES 786
Qy 113 -----CNCCNNLHHDIERFAKACLG-----RNPAFQPKIGKQ 149
Db 787 YLSNOEFPDPREPCNLG-----TCLGFGVTCGRPRCPPEPGCSHPLIPSGH 831
Qy 150 L-----GNVKP-----OHNKGCNCRSGCLKNYCEYEA 178
Db 832 CCPTQCGRYGVTTASGETLPDPLDPTCSLCTCQGREHODGEEFEGPAGSCWCRCQAG 891
Qy 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPTMOTGGLEGSH 219
Db 892 QVSCVRLQCPPLCKLQVTERGSCCPRCGLAHGEHP-----EGSR 934
Qy 220 YLPPTKFSGLPRFSDHRRSPSCISWEVEATCA---CL--LAQEEAEKHEHCKC 269
Db 935 WVPP-----DSACSSVCHEGV-VTCARIQCISSCAQPRQGHDCCPQC 977

RESULT 9
US-10-137-129A-3
; Sequence 3, Application US/10137129A
; Publication No. US20030124108A1
; GENERAL INFORMATION:
; APPLICANT: Frohman, Michael
; APPLICANT: Morris, Andrew
; APPLICANT: Engebrecht, Joanne
; FILE REFERENCE: Onyx2004-DIV2
; CURRENT FILING DATE: 2003-01-08
; CURRENT APPLICATION NUMBER: US/10/137.129A
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4588
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-137-129A-3

Query Match          6.8%; Score 109.5; DB 14; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.043;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

Qy 1 MVICOLKGGTQMLCIDNSRRELKALHLVPOYQDNNYLOSDVDPKPMALVGLFPLASTK 60
Db 436 IVTVQVLPQCECRCDQSRDRL--CH-----GKGFLECGICRCDTGYIGNKCECQTQ 486
Qy 61 LNLITQOLEGA-----LPSVVGNSAFPSGTLPGPPKITLAGYCDCEFA-----103
Db 487 -GRSSOLEGSCRDNNNSIICSLGDCVCGOCLCHTSDFPG--KLIYQYCECDTINCER 543
Qy 104 -SGDFCN--CNCNCCNNLHHDIERF-----KAICACLRNPEAFQPKIGKQLGN 152
Db 544 YNGQVCGPGGRGLCFCKGK--RCHPGFEGSACQERTTEGCL--NPRRVCS--GRGR---595
Qy 153 VKQHNKGCNCRSGCLKNYCEYEAQIM-----CSSIC-----KCIGCKNYEESP 198
Db 596 -----CRC-----NVCECHSGYQLPLQCEPCGCPSPCGKYISCAECLAFKEGP 638
```

RESULT 11
US-10-133-172-20
; Sequence 20, Application US/1013172
; Publication No. US20030104987A1
; GENERAL INFORMATION:
; APPLICANT: WEIGEL, PAUL H
; APPLICANT: WEIGEL, JANET A
; TITLE OF INVENTION: METHODS OF USING THE HYALURONAN RECEPTOR FOR ENDOCYTOSIS
; FILE REFERENCE: 5864.014
; CURRENT APPLICATION NUMBER: US/10/133,172
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,468
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 09/842,930
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 1653
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-172-20

Query Match 6.6%; Score 109; DB 15; Length 1653;
Best Local Similarity 26.9%; Pred. No. 0.13; Mismatches 18; Conservative 49; Indels 36; Gaps 12;
Matches 49; Conservative 18; Mismatches 18; Indels 36; Gaps 12;
QY 97 GYDCDFAS--GDPCNNC-----NCNNCCNNLHHDIERFAIKACLRNPEAFQPKIGKQL 150
DB 516 GSCDDVGRWGHCDNATDNCNGTCHTSANCLTNSDGTASC--KCAAGFQ---GNGTI 570
QY 151 G---NVKQHNKGC---NCRSGCLKNVCEG---YEAQ-IMCSSICKCI-----GC-KNY 194
DB 571 CTAINACEISNGCGSAKADCKRTTPGRVCTCKAGYTGDIIVCLINPCLENHGGCDKNA 630
QY 195 E---ESPERKTLMSMPNYMOTGGLGSHYLPPTKFSGLPRFSGHRRPSSCISWEYVEATC 251
DB 631 ECTQTGPNQAACNCLPAYTGDKVKCTLIYVCLTKNGCSEF-----AICNHTGQVERTC 684
QY 252 AC 253
DB 685 TC 686

RESULT 12
US-09-845-583-2
; Sequence 2, Application US/09845583
; Patent No. US20020142954A1
; GENERAL INFORMATION:
; APPLICANT: Burgeson, Robert
; APPLICANT: Brunken, William Joseph
; APPLICANT: Champilaud, Marie-France
; APPLICANT: Hunter, Dale
; TITLE OF INVENTION: LAMININ 15 AND USES THEREOF
; FILE REFERENCE: 10287-056001
; CURRENT APPLICATION NUMBER: US/09/845,583
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/200,863
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3635
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-845-583-2

Query Match 6.5%; Score 106.5; DB 10; Length 3635;
Best Local Similarity 21.9%; Pred. No. 0.68;
Matches 56; Conservative 26; Mismatches 87; Indels 87; Gaps 15;

QY 94 TLAGYCD-CFA-----SGDFCNCNCNNCCNNLHHDIERFAIKACLRNPE 139
DB 250 TCGGSCDRCCPGFNQOPWKPATTDSANECQSCNCHGAYDYDPEVDR-----RNAS 302
QY 140 AFQPKIGKQLGNV---KPOHNKGCNRRSGCLKNVCEYEAQIMCSSICKIGCKN---- 193
DB 303 QNQDNVYQG--GGVCLDCQHHTTGINCER--CLPGFFRAPDQPLDSPHYCRPCDCESDPT 358
QY 194 ---YEESPERKTLMSMPNYMOTGGL-----EG-----SHYLPPTKFSGLPRESH-DRRPSS 240
DB 359 DGTCEDLTGR--CYCRPNF--TGELCAACAEGYTDPPHCYP-----LPSFPHNDTREQV 408
QY 241 CISWEVVEATCACLQAQEEAEKE-----HCSKC----- 269
DB 409 LPAGQIVNCDNAAGTQGNACRDPRLGRVCVKPNFRGAHCELCAFGPHGSPCHPCQCSS 468
QY 270 --LAEQMILEEFGRL 283
DB 469 PGVANSLCDPESGQCM 484

RESULT 13

US-10-037-182-4
; Sequence 4, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thyboll, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3635
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-182-4

Query Match 6.5%; Score 106.5; DB 15; Length 3635;
Best Local Similarity 21.9%; Pred. No. 0.68; Mismatches 26; Indels 87; Gaps 15;
Matches 56; Conservative 26; Mismatches 26; Indels 87; Gaps 15;

QY 94 TLAGYCD-CFA-----SGDFCNCNCNNCCNNLHHDIERFAIKACLRNPE 139
DB 250 TCGGSCDRCCPGFNQOPWKPATTDSANECQSCNCHGAYDYDPEVDR-----RNAS 302
QY 140 AFQPKIGKQLGNV---KPOHNKGCNRRSGCLKNVCEYEAQIMCSSICKIGCKN---- 193
DB 303 QNQDNVYQG--GGVCLDCQHHTTGINCER--CLPGFFRAPDQPLDSPHYCRPCDCESDPT 358
QY 194 ---YEESPERKTLMSMPNYMOTGGL-----EG-----SHYLPPTKFSGLPRESH-DRRPSS 240
DB 359 DGTCEDLTGR--CYCRPNF--TGELCAACAEGYTDPPHCYP-----LPSFPHNDTREQV 408
QY 241 CISWEVVEATCACLQAQEEAEKE-----HCSKC----- 269
DB 409 LPAGQIVNCDNAAGTQGNACRDPRLGRVCVKPNFRGAHCELCAFGPHGSPCHPCQCSS 468
QY 270 --LAEQMILEEFGRL 283
DB 469 PGVANSLCDPESGQCM 484

RESULT 14

US-09-864-761-42967
; Sequence 42967, Application US/09864761

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:34:58 ; Search time 20.638 Seconds
(without alignments)
612.991 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQLHTFEKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109.5	6.6	769	US-08-789-078-1	Sequence 1, Appli
2	109.5	6.6	769	US-08-752-633-1	Sequence 1, Appli
3	109.5	6.6	769	US-08-476-062A-45	Sequence 45, Appl
4	109.5	6.6	769	US-07-728-215-31	Sequence 31, Appl
5	109.5	6.6	769	US-08-938-085A-31	Sequence 31, Appl
6	109.5	6.6	769	US-10-072-844-31	Sequence 31, Appl
7	109.5	6.6	769	PCR-US95-04886-1	Sequence 1, Appli
8	109.5	6.6	769	PCR-US96-01314-45	Sequence 45, Appl
9	108	6.6	676	US-08-630-172-10	Sequence 10, Appl
10	108	6.6	676	US-09-375-419-10	Sequence 10, Appl
11	108	6.6	739	US-08-444-818-148	Sequence 148, App
12	108	6.6	2995	US-08-444-818-138	Sequence 2, Appl
13	101	6.1	465	US-08-833-678A-2	Sequence 2, Appli
14	101	6.1	465	US-08-529-169A-2	Sequence 2, Appli
15	101	6.1	465	US-09-483-799-2	Sequence 2, Appli
16	101	6.1	590	US-08-850-328-3	Sequence 3, Appli
17	101	6.1	631	US-08-833-678A-1	Sequence 1, Appli
18	101	6.1	631	US-08-529-169A-1	Sequence 1, Appli
19	101	6.1	631	US-09-483-799-1	Sequence 1, Appli
20	101	6.1	1021	US-07-910-760-12	Sequence 12, Appl
21	101	6.1	1021	US-08-440-519-12	Sequence 12, Appl
22	101	6.1	1021	US-08-440-519-12	Sequence 12, Appl
23	101	6.1	1786	US-08-444-818-54	Sequence 54, Appl
24	101	6.1	2261	US-08-444-818-66	Sequence 66, Appl
25	101	6.1	2436	US-08-444-818-75	Sequence 75, Appl
26	101	6.1	2772	US-08-444-818-89	Sequence 89, Appl
27	101	6.1	2894	US-08-466-975A-23	Sequence 23, Appl

28	101	6.1	2894	2	US-08-391-671A-23	Sequence 23, Appl
29	101	6.1	2894	3	US-08-467-902A-23	Sequence 23, Appl
30	101	6.1	2894	3	US-09-275-265-23	Sequence 23, Appl
31	101	6.1	2894	4	US-09-941-611-23	Sequence 23, Appl
32	101	6.1	3011	1	US-08-440-103-36	Sequence 36, Appl
33	101	6.1	3011	1	US-08-440-542-36	Sequence 36, Appl
34	101	6.1	3011	1	US-07-910-760-10	Sequence 10, Appl
35	101	6.1	3011	1	US-08-440-519-10	Sequence 10, Appl
36	101	6.1	3011	1	US-08-231-368-36	Sequence 36, Appl
37	101	6.1	3011	1	US-08-440-210-36	Sequence 36, Appl
38	101	6.1	3011	3	US-09-388-874-2	Sequence 2, Appli
39	101	6.1	3011	4	US-09-046-604-36	Sequence 36, Appl
40	101	6.1	3011	4	US-08-440-549-10	Sequence 10, Appl
41	101	6.1	3011	4	US-08-850-328-1	Sequence 1, Appli
42	101	6.1	3011	4	US-09-916-359-2	Sequence 2, Appli
43	97.5	5.9	921	4	US-09-699-266A-9	Sequence 9, Appli
44	97.5	5.9	969	2	US-08-284-941-2	Sequence 2, Appli
45	97.5	5.9	969	2	US-08-447-642-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-789-078-1
; Sequence 1, Application US/08789078
; Patent No. 5843885
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; NUMBER OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,078
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,513
; FILING DATE: 19-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: 816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:

OTHER INFORMATION: /note= "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
AUTHORS: Power,
TITLE: LFA-1 Amino acid sequence (B2) (from human
TITLE: tonsil)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
US-08-752-633-1

Query Match 6.6%; Score 109.5; DB 2: Length 769;
Best Local Similarity 22.3%; Pred. No. 0.042;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MWICQLKGGTQMLCTDNRSTRELKALHLVPOYQDQNNYLSQDVPPKMTALVGRFLPASTK 60
DB 436 IYTVQVLPOCECRCDQSRDSL--CH-----GKGFLECGICRCDTGYIGKNECQQTQ 486
QY 61 LNLITQQLGGA-----LPSVNVGSAPFSGSTLPGPPKITLAGYCDCEFA----- 103
DB 487 -GRSSELEGGCRKDNNSIICSGLDGCVGCQCLHTSDVPG--KLIYGOYCECDTINCER 543
QY 104 -SGDFCN-----CNCNCCNNLHHDIERF-----KAIKACLGRLNPEAFQPKIGKQGLGN 152
DB 544 YNGQVCGGPGRLGFCGKC--RCHPFGEGSACQERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VKPQHNKGCNCRSSCLKNYCEYEAQIM-----CSSIC-----KCI GKKNYESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCQECPCGSPCKYISCAECLKFEKGP 638

RESULT 3
US-08-476-062A-45
Sequence 45, Application US/08476062A
Patent No. 5877275
GENERAL INFORMATION:
APPLICANT: Arnaout, M. Amin
TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,062A
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,081
FILING DATE: 21-MAR-1994
APPLICATION NUMBER: 07/637,830
FILING DATE: 04-JAN-1991
APPLICATION NUMBER: 07/539,842
FILING DATE: 18-JUN-1990
APPLICATION NUMBER: 07/212,573
FILING DATE: 28-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/068003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906

TELEX: 200154
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-476-062A-45
Query Match 6.6%; Score 109.5; DB 2: Length 769;
Best Local Similarity 22.3%; Pred. No. 0.042;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MWICQLKGGTQMLCTDNRSTRELKALHLVPOYQDQNNYLSQDVPPKMTALVGRFLPASTK 60
DB 436 IYTVQVLPOCECRCDQSRDSL--CH-----GKGFLECGICRCDTGYIGKNECQQTQ 486
QY 61 LNLITQQLGGA-----LPSVNVGSAPFSGSTLPGPPKITLAGYCDCEFA----- 103
DB 487 -GRSSELEGGCRKDNNSIICSGLDGCVGCQCLHTSDVPG--KLIYGOYCECDTINCER 543
QY 104 -SGDFCN-----CNCNCCNNLHHDIERF-----KAIKACLGRLNPEAFQPKIGKQGLGN 152
DB 544 YNGQVCGGPGRLGFCGKC--RCHPFGEGSACQERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VKPQHNKGCNCRSSCLKNYCEYEAQIM-----CSSIC-----KCI GKKNYESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCQECPCGSPCKYISCAECLKFEKGP 638

RESULT 4
US-07-728-215-31
Sequence 31, Application US/07728215
Patent No. 5962643
GENERAL INFORMATION:
APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Bruesgemann & Clark
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/728,215
FILING DATE: 19910711
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P31 8717
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-728-215-31

Db 487 -GRSSQELGSGCRKDNNSIICSLGDCVCGQCCLHTSDVPG--KLIYGOYCECDTINCER 543
QY 104 -SGDFCENN-----CNCNCCNNLHHDIERF-----KAIKACLGSRNPEAFQPKIGKQLGN 152
Db 544 YNQVCGGPGRLCFCGKC--RCHPGFEGSACQCERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VRPQHNKGCNRRSGCLKNYCYEYQAQIM-----CSSIC-----KCI GKCKNYEESP 198
Db 596 -----CRC-----NVCECHSGYQLPLCQECPCSPCGKYISCAECLKFEKGP 638

RESULT 7

PCT-US95-04886-1
; Sequence 1, Application PC/TUS9504886
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Slahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: 816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..22
; OTHER INFORMATION: /label= signal
; OTHER INFORMATION: /note= "signal sequence"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 449..496
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:

; NAME/KEY: Region
; LOCATION: 541..581
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 582..617
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 701..723
; OTHER INFORMATION: /label= trans
; OTHER INFORMATION: /note= "transmembrane domain"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 724..769
; OTHER INFORMATION: /label= cyto
; OTHER INFORMATION: /note= "cytoplasmic domain"
; PUBLICATION INFORMATION:
; AUTHORS: Pigott,
; TITLE: LFA-1 Amino acid sequence (B2) (from human
; TITLE: tonsil)
; JOURNAL: The Adhesion Molecule Facts Book
; PAGES: 96-96
; DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
PCT-US95-04886-1

Query Match 6.6%; Score 109.5; DB 5; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.042;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVICOLKGTQMLCINDNSTREKALHLVPQYODONNYLOSDVPKPM TALVGRFLPASTK 60
Db 436 IVTQVLPQCECRCDQSRDSL--CH-----GKGFLECGICRCDTGYIGNKCBQQTQ 486
QY 61 LNLITQOLEGA-----LPSVNVGSAPPSGSLPGPKITLACYCCDCA----- 103
Db 487 -GRSSQELGSGCRKDNNSIICSLGDCVCGQCCLHTSDVPG--KLIYGOYCECDTINCER 543
QY 104 -SGDFCENN-----CNCNCCNNLHHDIERF-----KAIKACLGSRNPEAFQPKIGKQLGN 152
Db 544 YNQVCGGPGRLCFCGKC--RCHPGFEGSACQCERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VRPQHNKGCNRRSGCLKNYCYEYQAQIM-----CSSIC-----KCI GKCKNYEESP 198
Db 596 -----CRC-----NVCECHSGYQLPLCQECPCSPCGKYISCAECLKFEKGP 638

RESULT 8

PCT-US96-01314-45
; Sequence 45, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaut
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/01314

```

; FILING DATE: 30-JAN-96
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/380,167
; FILING DATE: 30-JAN-95
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00786/267001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US96-01314-45

Query Match
Best Local Similarity 22.3%; Score 109.5; DB 5; Length 769;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDNNYLSQSDVPKPMYALVGRFLPASTK 60
Db 436 IVTVQVLPOCECRCDQSRDRSL--CH-----GKGFLEGGICRCDTGYIGKNCCECOTQ 486

QY 61 LNLITQOLEGA-----LPSVVGSAFPGSGTLPGPKITLAGYCDCA----- 103
Db 487 -GRSSQLEGGSCRKDNNSIICSGGLGDCVCGQCLCHTSVDPG--KLIYQYCECDTINCER 543

QY 104 -SGDFCINN-----CNCNCCNHLHDIERF-----KAIKACGLGRNPEAFQPKIGKGLGN 152
Db 544 YNGQVCGGPGRLCFEGKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR--- 595

QY 153 VKPOHKGNCNRRSGCLKNYCEYEAQIM-----CSSIC-CKIGCKNYEESP 198
Db 596 -----CRC-----NVCECHSGYQLPLCQECPCGPGSPCGYISCAECLKFEKGP 638

RESULT 9
US-08-630-172-10
; Sequence 10, Application US/08630172
; Patent No. 6060054
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYMPHOCYTE VETO
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,172
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 676 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-630-172-10

Query Match
Best Local Similarity 21.5%; Score 108; DB 3; Length 676;
Matches 51; Conservative 26; Mismatches 86; Indels 74; Gaps 14;

QY 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDNNYLSQSDVPKPMYALVGRFLPASTK 60
Db 414 IVTVQVLPOCECRCDQSRDRSL--CH-----GKGFLEGGICRCDTGYIGKNCCECOTQ 464

QY 61 LNLITQOLEGA-----LPSVVGSAFPGSGTLPGPKITLAGYCDCA----- 103
Db 465 -GRSSQLEGGSCRKDNNSIICSGGLGDCVCGQCLCHTSVDPG--KLIYQYCECDTINCER 521

QY 104 -SGDFCINN-----CNCNCCNHLHDIERF-----KAIKACGLGRNPEAFQPKIGKGLGN 152
Db 522 YNGQVCGGPGRLCFEGKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR--- 573

QY 153 VKPOHKGNCNRRSGCLKNYCEYEAQIM-----CSSICKICIGCKNYEESP 198
Db 574 -----CRC-----NVCECHSGYQLPLCQECPCGPGSPCGYISCAECLKFEKGP 615

RESULT 10
US-09-375-419-10
; Sequence 10, Application US/09375419
; Patent No. 6264950
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYMPHOCYTE VETO
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/375,419
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/630,172
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 676 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-375-419-10

Query Match
6.6%; Score 108; DB 3; Length 676;

```

288	QD	AVSRTOR-----RGRTRGKPGGINRFVAPCRSPGMFDSVLCGYDA-----	332
187	KI	CGCKNVEESPERRKTLMSMPNMYMOTGGL-----EG-----	SHVLPPTKFS 227
333	--GC	AWYIELLP-AETTVRURAYANTPGLPVCDHLEFWEGVTGLTHDAHFLSQTOKS 388	
228	G--LP	FSHRRRSPSCISWEVEVATCACLLAQGEAEAKEHCSCKL 270	
389	GENLP-----	YLVAQNTVCAARAQAPPPSWDQMKCL 420	

RESULT 12
US-08-444-818-138
; Sequence 138, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.

```

NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444.818

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 2995 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-138

131	QY	KACLGRLNPEAFQPKIGKGLGNGVNRPOINK	---GCNCRRSGLCKN	---YCEYEAQIMC	SSIC	181
132	QY					182
133	QY					183
134	QY					184
135	QY					185
136	QY					186
137	QY					187
138	QY					188
139	QY					189
140	QY					190
141	QY					191
142	QY					192
143	QY					193
144	QY					194
145	QY					195
146	QY					196
147	QY					197
148	QY					198
149	QY					199
150	QY					200
151	QY					201
152	QY					202
153	QY					203
154	QY					204
155	QY					205
156	QY					206
157	QY					207
158	QY					208
159	QY					209
160	QY					210
161	QY					211
162	QY					212
163	QY					213
164	QY					214
165	QY					215
166	QY					216
167	QY					217
168	QY					218
169	QY					219
170	QY					220
171	QY					221
172	QY					222
173	QY					223
174	QY					224
175	QY					225
176	QY					226
177	QY					227
178	QY					228
179	QY					229
180	QY					230
181	QY					231
182	QY					232
183	QY					233
184	QY					234
185	QY					235
186	QY					236
187	QY					237
188	QY					238
189	QY					239
190	QY					240
191	QY					241
192	QY					242
193	QY					243
194	QY					244
195	QY					245
196	QY					246
197	QY					247
198	QY					248
199	QY					249
200	QY					250
201	QY					251
202	QY</					

RESULT 13

US-08-833-678A-2
; Sequence 2, Application US/08833678A
; Patent No. 5989905
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833.678A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/529,169
; FILING DATE: 15-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-833-678A-2

Query Match 6.1%; Score 101; DB 2; Length 465;
Best Local Similarity 24.8%; Pred. No. 0.14;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;
Qy 79 SAPP-SGSTLPGPPIKTLGACGDCFCASGDFCNCCNCCNNL-----HHDIREFKAI 130
Db 232 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNTCTVTQVDFSLDPTFTTITLTP 286
Qy 131 KACIGRNPEAFQPKIGKGLG---NVRPQHNGKNCRRSGCLKN--YCEGYEAGIMCSSI 185
Db 287 QDAVSRTQR--RGRTGKGKGIYRFVAPGE-----RPSGMFSDSVLCECYDA----- 331
Qy 186 KCICGCKNYEESPERKTLMSNPYMTQGL-----EG-----SHYLPPTKF 226
Db 332 ----GCAYWELTP-AETTVRLRAYMNTPLPVCODHLEFWEVFTGLTHIDAHFLSQTQ 386
Qy 227 SG--LPFRSHDRPSSCISWEVVEATCACLLAQGEAEKEHCSCKL 270
Db 387 SGENLP-----YLVAYQATVCAQAQAPPPSQDMWKCL 419

RESULT 14

US-08-529-169A-2

; Sequence 2, Application US/08529169A
; Patent No. 6194140
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/529,169A
; FILING DATE: 15-SEP-1995
; CLASSIFICATION: 4325
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-529-169A-2

Query Match 6.1%; Score 101; DB 3; Length 465;
Best Local Similarity 24.8%; Pred. No. 0.14;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;
Qy 79 SAPP-SGSTLPGPPIKTLGACGDCFCASGDFCNCCNCCNNL-----HHDIREFKAI 130
Db 232 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNTCTVTQVDFSLDPTFTTITLTP 286
Qy 131 KACIGRNPEAFQPKIGKGLG---NVRPQHNGKNCRRSGCLKN--YCEGYEAGIMCSSI 185
Db 287 QDAVSRTQR--RGRTGKGKGIYRFVAPGE-----RPSGMFSDSVLCECYDA----- 331
Qy 186 KCICGCKNYEESPERKTLMSNPYMTQGL-----EG-----SHYLPPTKF 226
Db 332 ----GCAYWELTP-AETTVRLRAYMNTPLPVCODHLEFWEVFTGLTHIDAHFLSQTQ 386
Qy 227 SG--LPFRSHDRPSSCISWEVVEATCACLLAQGEAEKEHCSCKL 270
Db 387 SGENLP-----YLVAYQATVCAQAQAPPPSQDMWKCL 419

RESULT 15

US-09-483-799-2
; Sequence 2, Application US/09483799
; Patent No. 6472180
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO

Search completed: July 24, 2003, 13:49:27
Job time : 21.638 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 13:22:12 ; Search time 59.9007 Seconds
(without alignments)
792.299 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRLSQLHTEFKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	100.0	299	21	AA198464 Human testis speci
2	1351	76.0	295	21	AA198463 Mouse testis speci
3	667.5	40.5	528	24	ABG74678 Human CGPD protein
4	470	28.5	950	22	ABB62035 Drosophila melanog
5	446	27.1	147	22	ABB96025 Human testicular a
6	446	27.1	147	22	AA195330 Human reproductive
7	441	26.8	438	20	AAW83392 Caenorhabditis ele
8	378	23.0	280	22	ABG17958 Novel human diagno
9	278	16.9	53	22	AA1976158 Human colon cancer

10	247.5	15.0	609	24	ABP70527 Histone deacetylase
11	216	13.1	251	22	AA193348 Human polypeptide
12	200	12.1	243	22	ABB68888 Drosophila melanog
13	113	6.9	1981	22	ABB61657 Human kiellin-like
14	112.5	6.8	1192	24	ABP97376 Human kiellin-like
15	112.5	6.8	1207	24	ABP97377 Human kiellin-like
16	112.5	6.8	1477	24	ABP97371 Human kiellin-like
17	112.5	6.8	1535	24	ABP97374 Human kiellin-like
18	112.5	6.8	1593	24	ABP97369 Human kiellin-like
19	112	6.8	1774	23	ABG69800 Human REMAP-6 prot
20	109.5	6.6	699	20	AAW81840 Human LFA-1 beta c
21	109.5	6.6	699	24	ABU04052 Human expressed pr
22	109.5	6.6	761	24	ABU04045 Human expressed pr
23	109.5	6.6	769	9	ABP80836 Beta subunit of hu
24	109.5	6.6	769	11	AA197113 Recombinant beta-s
25	109.5	6.6	769	16	AA1980108 LFA-1 beta subunit
26	109.5	6.6	769	21	AA193974 Human CD18. Homo
27	109.5	6.6	769	23	ABP54798 Human expressed pr
28	109.5	6.6	769	24	ABU04042 Human expressed pr
29	109.5	6.6	769	24	ABU04044 Human expressed pr
30	109.5	6.6	769	24	ABU04046 Human expressed pr
31	109.5	6.6	769	24	ABU04047 Human expressed pr
32	109.5	6.6	769	24	ABU04048 Human expressed pr
33	109.5	6.6	769	24	ABU04049 Human expressed pr
34	109.5	6.6	769	24	ABU04050 Human expressed pr
35	109.5	6.6	769	24	ABU04053 Human expressed pr
36	109.5	6.6	769	24	ABU04054 Human expressed pr
37	109.5	6.6	793	22	ABG16399 Novel human diagno
38	109	6.6	1653	24	ABG72514 Human 190kDa Hyalu
39	108.5	6.6	769	13	AA1924256 Beta-subunit CD18
40	108	6.6	676	18	AAW35853 Human CD18 for use
41	108	6.6	676	24	ABU04051 Human expressed pr
42	108	6.6	3011	13	AA1921519 Compiled HCV seque
43	106.5	6.5	3635	23	ABB81589 Mouse laminin alph
44	106.5	6.5	3635	23	AAW50357 Mouse laminin-15 a
45	105.5	6.4	695	22	ABG51014 Human liver peptid

ALIGNMENTS

RESULT 1
AA198464
ID AA198464 standard; Protein; 299 AA.
XX
AC AA198464;
XX
XX
DT 25-APR-2000 (first entry)
XX
DE Human testis specific factor tesmin SEQ ID NO:5.
XX
DE Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
OS Homo sapiens.
XX
PN WO200004147-A1.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-JP03859.
XX
PR 17-JUL-1998; 98JP-0219856.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
DR WPI; 2000-147785/13.
DR N-PSDB; AA288156.
XX

PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 useful for investigation of germ cell differentiation and sterility -
 Claim 1; Page 50-52; 63pp; Japanese.

XX The present sequence represents a male germ cell regulatory factor
 expressed specifically in spermatocytes, designated tesmin. Tesmin
 can be used in the investigation of the mechanisms of germ cell
 differentiation and sterility.

XX Sequence 299 AA;

Query Match 100.0%; Score 1647; DB 21; Length 299;
 Best Local Similarity 100.0%; Pred. No. 4.3e-141;
 Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MVICOLKGGTOMLCIDNSRTRRELKALHLVPOYQDNNYQSDVDPKMTALVGRFLPASTK 60
 Db 1 MVICOLKGGTOMLCIDNSRTRRELKALHLVPOYQDNNYQSDVDPKMTALVGRFLPASTK 60

Oy 61 LNLITQLEGALPSVWNGSAFPPSGSTLPGPKITLAGYCDGFCNCCNCCNNL 120
 Db 61 LNLITQLEGALPSVWNGSAFPPSGSTLPGPKITLAGYCDGFCNCCNCCNNL 120

Oy 121 HHDIERFKAICACLRNPEAFQPKIGKQGLGNVPOHNGKNCRRSGCLKNYCEYEAQI 180
 Db 121 HHDIERFKAICACLRNPEAFQPKIGKQGLGNVPOHNGKNCRRSGCLKNYCEYEAQI 180

Oy 181 MCSSTICKICGCKNYEESPERKTLMPNMYMTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240
 Db 181 MCSSTICKICGCKNYEESPERKTLMPNMYMTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240

Oy 241 CISWEVVEATCACLAAQGEAEKEHCSCKLAEOQMLEEFGRLCSQILHTEFKSKGLKME 299
 Db 241 CISWEVVEATCACLAAQGEAEKEHCSCKLAEOQMLEEFGRLCSQILHTEFKSKGLKME 299

RESULT 2
 AAY68463
 ID AAY68463 standard; Protein; 295 AA.

XX AC AAY68463;

XX 25-APR-2000 (first entry)

DE Mouse testis specific factor tesmin SEQ ID NO:4.

XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 differentiation regulatory factor; male germ cell regulatory actor;
 germ cell differentiation; sterility.

XX Mus musculus.

XX WO200004147-A1.

XX 27-JAN-2000.

XX 16-JUL-1999; 95WO-JP03859.

XX 17-JUL-1998; 98JP-0219856.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX WPI; 2000-147785/13.

DR N-PSDB; AAZ88155, AAZ88157.

XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 useful for investigation of germ cell differentiation and sterility -
 Claim 1; Page 47-49; 63pp; Japanese.

XX The present sequence represents a male germ cell regulatory factor
 expressed specifically in spermatocytes, designated tesmin. Tesmin
 can be used in the investigation of the mechanisms of germ cell
 differentiation and sterility.

XX Sequence 295 AA;

Query Match 76.0%; Score 1251; DB 21; Length 295;
 Best Local Similarity 76.3%; Pred. No. 3.6e-105;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Oy 1 MVICOLKGGTOMLCIDNSRTRRELKALHLVPOYQDNNYQSDVDPKMTALVGRFLPASTK 60
 Db 1 MVICOLKGGTOMLCIDNSRTRRELKALHLVPOYQDNNYQSDVDPKMTALVGRFLPASTK 60

Oy 61 LNLITQLEGALPSVWNGSAFPPSGSTLPGPKITLAGYCDGFCNCCNCCNNL 120
 Db 61 LNLITQLEGALPSVWNGSAFPPSGSTLPGPKITLAGYCDGFCNCCNCCNNL 120

Oy 121 HHDIERFKAICACLRNPEAFQPKIGKQGLGNVPOHNGKNCRRSGCLKNYCEYEAQI 180
 Db 121 HHDIERFKAICACLRNPEAFQPKIGKQGLGNVPOHNGKNCRRSGCLKNYCEYEAQI 180

Oy 181 MCSSTICKICGCKNYEESPERKTLMPNMYMTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240
 Db 181 MCSSTICKICGCKNYEESPERKTLMPNMYMTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240

Oy 241 CISWEVVEATCACLAAQGEAEKEHCSCKLAEOQMLEEFGRLCSQILHTEFKSKGLKME 299
 Db 241 CISWEVVEATCACLAAQGEAEKEHCSCKLAEOQMLEEFGRLCSQILHTEFKSKGLKME 299

RESULT 3
 ABG74678
 ID ABG74678 standard; Protein; 528 AA.

XX AC ABG74678;

XX 10-MAY-2003 (first entry)

DE Human CGDD protein 2369279CD1 SEQ ID 4.

XX Human; cell growth, differentiation and death; CGDD; cardiant; cancer;
 cytostatic; neuroprotective; gene therapy; diagnosis; cardiovascular;
 neurological disorder.

XX Homo sapiens.

XX WO2003014322-A2.

XX 20-FEB-2003.

XX 08-AUG-2002; 2002WO-US25465.

XX 08-AUG-2001; 2001US-311017P.

XX 17-AUG-2001; 2001US-313070P.

XX 17-AUG-2001; 2001US-313071P.

XX 24-AUG-2001; 2001US-314678P.

XX 31-AUG-2001; 2001US-316692P.

XX 07-SEP-2001; 2001US-317913P.

XX 14-SEP-2001; 2001US-322182P.

XX 07-DEC-2001; 2001US-340747P.

XX 20-DEC-2001; 2001US-342761P.

XX 29-MAR-2002; 2002US-369129P.

XX (INCY-) INCYTE GENOMICS INC.

XX Azimzai Y, Barroso I, Baughn MR, Becha SD, Borowsky ML, Duggan BM;
 PI Elliott VS, Emerling BM, Forsythe IJ, Gietzen KJ, Gorvad AE;
 PI Graul RC, Griffin JA, Gururajan R, Hafalia AJA, Ison CH, Kable AE;
 PI Khan FA, Lee S, Lee SY, Li JX, Reddy R, Richardson TW;
 PI Sprague WW, Swarnakar A, Tang YT, Warren BA, Xu Y, Yao MG, Yue H;

XX OS Homo sapiens.
XX PN WO200155317-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01329.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483232/52.
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer
XX Claim 11; SEQ ID NO 1409; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a protein of the
CC invention.
XX
SQ Sequence 147 AA;
Query Match 27.1%; Score 446; DB 22; Length 147;
Best Local Similarity 98.9%; Pred No. 1.7e-32;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MYICQLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYQSDVPKMTALVGRFLPASTK 60
DQ 60 MYICQLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYQSDVPKMTALVGRFLPASTK 119
QY 61 LNLITQOLEGALPSVYVNGSAPFSGSTLP 88
DQ 120 LNLITQOLEGALPSVYVNGSAPFSGSTXP 147
RESULT 6
AAW95330
ID AAW95330 standard; Protein: 147 AA.
XX
AC AAW95330;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen SEQ ID NO: 3988.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.

Db 251 LHQKCHCKSKCLNKECYEAKVPCDTRCKCKQNTETRYMTRYNKSGGAVSNTNAL 310
QY 204 MSMPNYMOT-----GGD-----EGSHYLPPTKFSGLPRESHRRPSSCISW-----EVV 247
Db 311 MSFTNASSTATPDSGSGVVDTEHGGDY-EDMLLSHKPKAVMDPRP---FPWYMTDEVV 366
QY 248 EATCACLAAQGEA-----EKEHCKSLAEOMILEEFGRCLSQIL 287
Db 367 EAATCMVAQAEBEALNKEKVQTEDEKLNMEKLVLRFRGRCLEQMI 412

RESULT 8
ABG17958
ID ABG17958 standard; Protein; 280 AA.
AC ABG17958;
XX
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17949.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS82145.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 48317; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 280 AA;

Query Match 23.0%; Score 378; DB 22; Length 280;
Best Local Similarity 41.3%; Pred. No. 5.5e-26;
Matches 78; Conservative 23; Mismatches 54; Indels 34; Gaps 5;

QY 98 YDCDFASGDFCNCCNCCNCCNLDHDIERFKAICACLRNPEAFQPKIGKQGLGNVYKPOH 157
Db 114 YDCDFANGFCCNCCNCCNCCNLDHDIERFKAICACLRNPEAFQPKIGKQGLGNVYKPOH 173
QY 158 NKGCCNRRSGCLNKECYEAOIMCSCICKCKNKEEYSPERKTLMSMPNYMOTGLEG 217
Db 174 SKGCNKRSGCLNKECYEAOIMCSCICKCKNKEEYSPERKTLMSMPNYMOTGLEG 218
QY 218 SHYLPPTKFSGLPRF-----SHDRRPS-----SCISWEVVEATCACLAAQGEAEKEH 265
Db 219 -----OLEALKAFVEALIESVESRTSGMTQSCALQHTDLECSSIYGLKALRNHL 271
QY 266 CSKCLAEQM 274
Db 272 YQTRTAQL 280

RESULT 9
AAG76158
ID AAG76158 standard; Protein; 53 AA.
AC AAG76158;
XX
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6922.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 11.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
XX N-PSDB; AAH35563.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 11; Page 8365; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the

```
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 53 AA;
    Query Match          16.9%; Score 278; DB 22; Length 53;
    Best Local Similarity 100.0%; Pred. No. 8.2e-18;
    Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 VEATCACLLAAGEAEAEKHCSCLAEQMILEEFGRCLSQILHTEFKSKGLKME 299
Db 1 VEATCACLLAAGEAEAEKHCSCLAEQMILEEFGRCLSQILHTEFKSKGLKME 53

RESULT 10
ABP70527
ID ABP70527 standard; Protein; 609 AA.
XX
AC ABP70527;
XX
DT 22-APR-2003 (first entry)
XX
DE Histone deacetylase chromatin-associated protein complex subunit.
XX
KW Histone acetyltransferase; histone deacetylase; gene expression profile;
KW chromatin-associated protein; gene expression.
XX
OS Arabidopsis thaliana.
XX
FN WO2003000715-A1.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US19750.
XX
PR 22-JUN-2001; 2001US-300135P.
XX
PA (CERE-) CERES INC.
XX
XX Dang V, Okamuro J;
XX
DR WPI; 2003-175280/17.
XX
PT New chimeric polypeptide comprising a histone acetyltransferase
PT polypeptide segment and a segment comprising a histone deacetylase
PT chromatin-associated protein complex subunit, useful for modulating
PT gene expression in cells -
XX
PS Disclosure; Page 61; 85pp; English.
XX
CC The specification describes chimeric histone acetyltransferase
CC polypeptides. The chimeric polypeptides comprise a polypeptide segment
CC that exhibits histone acetyltransferase activity, and a polypeptide
CC segment having 40% or greater sequence identity to a subunit of a
CC histone deacetylase chromatin-associated protein complex. The chimeric
CC polypeptides are useful for determining gene expression profiles in
CC specific cells, for modulating gene expression in specific cells, and
CC for making genetically modified eukaryotes. The present sequence
CC represents an exemplary histone deacetylase chromatin-associated protein
CC complex subunit, which can be used to produce chimeric polypeptides of
CC the invention.
XX
SQ Sequence 609 AA;
    Query Match          15.0%; Score 247.5; DB 24; Length 609;
    Best Local Similarity 38.6%; Pred. No. 1e-13;
    Matches 49; Conservative 20; Mismatches 35; Indels 23; Gaps 6;

Qy 98 YCDCFASGDFC-NNCNCNNCCNN-LHHDIERFKAICLGNPEAFQPKICK----- 147
Db 341 YCEFAAGFYCIEPCSCINCNPDKHDKV-LATRKQIESRNPFLAFAPKVIIRNSDSIEV 399
```

```
Qy 148 GOLGNVKP---QHNKGNCRRSGCLKNYCEYEAQIMCISICKIGCKN-----YE 195
Db 400 GEDASKTPASARHRRKGCNCKNCKLKYCEYQGVGSGINCRCGCKNAFRKDGSLFE 459

Qy 196 ESPERKT 202
Db 460 QDEENET 466

RESULT 11
AAM93348
ID AAM93348 standard; Protein; 251 AA.
XX
AC AAM93348;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2895.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
FN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94268.
XX
830 Primers useful for synthesizing full length cDNA clones and their
use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 251 AA;
    Query Match          13.1%; Score 216; DB 22; Length 251;
    Best Local Similarity 100.0%; Pred. No. 2.4e-11;
    Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVICQLKGGTQMLCIDNSRTRRELKALHLVPOYQDNVLYQS 41
Db 210 MVICQLKGGTQMLCIDNSRTRRELKALHLVPOYQDNVLYQS 250

RESULT 12
ABB6888
```



```

AC ABP97376;
XX
DT 10-MAY-2003 (first entry)
XX
DE Human kielin-like protein NHP9, SEQ ID NO:18.
XX
KW Human; kielin-like protein; NHP; chromosome 7; development;
KW signal transduction; kielin-related disorder; diagnosis; treatment;
KW drug screening; clinical trial monitoring; cosmetic; nutraceutical.
XX
OS Homo sapiens.
XX
PN WO2003004609-A2.
XX
PD 16-JAN-2003.
XX
PF 03-JUL-2002; 2002WO-US21118.
XX
PR 03-JUL-2001; 2001US-302949P.
PR 29-AUG-2001; 2001US-315634P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Walke DW, Scoville J, Turner CA;
DR WPI; 2003-210352/20.
DR N-PSDB; ABZ75047.
XX
PT New human kielin-like nucleic acids, useful for preparing a composition
PT for treating human kielin-related disorders -
XX
PS Claim 3; Page 71-74; 85pp; English.
XX
CC The invention relates to novel human kielin-like proteins (referred to as
CC NHPs; ABP97376-ABP97379) and cDNAs encoding them (ABZ75039-ABZ75050). The
CC invention also encompasses recombinant expression vectors and host cells
CC comprising a nucleic acid of the invention. The NHP cDNAs were obtained
CC by aligning cDNAs from human kidney, foetal kidney, prostate and lymph
CC node mRNAs. The NHPs of the invention are all apparently encoded on
CC chromosome 7 and share sequence similarity with animal kielin proteins.
CC Kielins are secreted proteins implicated in a number of biological
CC processes such as development and signal transduction. The NHP nucleic
CC acids and proteins may be used in the diagnosis and treatment of human
CC kielin-related disorders, and may also be used in drug screening and
CC clinical trial monitoring. They may further be used in cosmetic and
CC nutraceutical applications. The present sequence represents an NHP
CC of the invention.
XX
SQ Sequence 1192 AA;
Query Match 6.8%; Score 112.5; DB 24; Length 1192;
Best Local Similarity 20.0%; Pred. No. 0.41;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;
QY 78 GSAPSGSTLPGP-----PKITLGYCDGCFASGDFCNCCN----- 112
DB 326 GKFAFGRRFPSPPTAAACHLCWEGSVCEPKACAPALCPFPARGDCPCDGCGEYLGES 385
QY 113 -----CNCCNNLHHDIERFKAICAG-----RNPEAFQPKIGKQ 149
DB 386 YLSNQFPDPREPNCNL-----TCLGGFVTCGRRPCEPPGCGSHLIPSGH 430
QY 150 L-----GNVXP-----QHNKGCNRRSGCLKNYCEYEA 178
DB 431 CCPTCGQCRYHGVTTASGETLPDLPDPTCSLCTCGREHQDEEFGPAGSCWCRCQAG 490
QY 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPNYMOTGLESH 219
DB 491 QVSCVRLQCPPLPKLQVTERGSCPCRCRGLAHGEHP-----EGSR 533
QY 220 YLPPTKFSGLPRFSDRRPSSCISWEVVEATCA---CL--LAOGEEAEKHCSC 269
DB 534 WPPP-----DSACSSCVCHGV-VTCARIQCISSCAQPRQGHDCPCQC 576

```

```

RESULT 15
ABP97377
ID ABP97377 standard; Protein; 1207 AA.
XX
AC ABP97377;
XX
DT 10-MAY-2003 (first entry)
XX
DE Human kielin-like protein NHP10, SEQ ID NO:20.
XX
KW Human; kielin-like protein; NHP; chromosome 7; development;
KW signal transduction; kielin-related disorder; diagnosis; treatment;
KW drug screening; clinical trial monitoring; cosmetic; nutraceutical.
XX
OS Homo sapiens.
XX
PN WO2003004609-A2.
XX
PD 16-JAN-2003.
XX
PF 03-JUL-2002; 2002WO-US21118.
XX
PR 03-JUL-2001; 2001US-302949P.
PR 29-AUG-2001; 2001US-315634P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Walke DW, Scoville J, Turner CA;
DR WPI; 2003-210352/20.
DR N-PSDB; ABZ75048.
XX
PT New human kielin-like nucleic acids, useful for preparing a composition
PT for treating human kielin-related disorders -
XX
PS Claim 3; Page 75-77; 85pp; English.
XX
CC The invention relates to novel human kielin-like proteins (referred to as
CC NHPs; ABP97378-ABP97379) and cDNAs encoding them (ABZ75039-ABZ75050). The
CC invention also encompasses recombinant expression vectors and host cells
CC comprising a nucleic acid of the invention. The NHP cDNAs were obtained
CC by aligning cDNAs from human kidney, foetal kidney, prostate and lymph
CC node mRNAs. The NHPs of the invention are all apparently encoded on
CC chromosome 7 and share sequence similarity with animal kielin proteins.
CC Kielins are secreted proteins implicated in a number of biological
CC processes such as development and signal transduction. The NHP nucleic
CC acids and proteins may be used in the diagnosis and treatment of human
CC kielin-related disorders, and may also be used in drug screening and
CC clinical trial monitoring. They may further be used in cosmetic and
CC nutraceutical applications. The present sequence represents an NHP
CC of the invention.
XX
SQ Sequence 1207 AA;
Query Match 6.8%; Score 112.5; DB 24; Length 1207;
Best Local Similarity 20.0%; Pred. No. 0.41;
Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;
QY 78 GSAPSGSTLPGP-----PKITLGYCDGCFASGDFCNCCN----- 112
DB 341 GKFAFGRRFPSPPTAAACHLCWEGSVCEPKACAPALCPFPARGDCPCDGCGEYLGES 400
QY 113 -----CNCCNNLHHDIERFKAICAG-----RNPEAFQPKIGKQ 149
DB 401 YLSNQFPDPREPNCNL-----TCLGGFVTCGRRPCEPPGCGSHLIPSGH 445
QY 150 L-----GNVXP-----QHNKGCNRRSGCLKNYCEYEA 178
DB 446 CCPTCGQCRYHGVTTASGETLPDLPDPTCSLCTCGREHQDEEFGPAGSCWCRCQAG 505
QY 179 QIMC-----SSIC-KCIGCKNY-EESPERKTLMSMPNYMOTGLESH 219

```

Db 506 QVSCVRLQCPPLPCKLQVTERGCCPRCRGCLANGGEHD-----EGSR 548
Qy 220 YLPPTKFSGLPRFSDRRPSSCISWEVVEATCA---CL--LAQGEAEKEHCKC 269
Db 549 WVFP-----DSACSCVCHEGV-VTCARTQCISSCAOPROGPHDCCPQC 591

Search completed: July 24, 2003, 13:36:58
Job time : 60.9007 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:48:49 ; Search time 191.573 Seconds

(without alignments)
6037.327 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 tatctgtggttgcccg.....aaaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL-frame+n2p,model -DEV-xlp

-Q/cgn2_1/USPTO_SPOOL/US0974323/runat_24072003_113104_17672/app_query.fasta_1.4750
-DB-SPREMBL_23 -QWMT-fascan -SUFFIX-rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE-pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE-LOCAL
-OUTFMT-ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER-US0974323@cgn.1.1.612,runat_24072003_113104_17672 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	No.	Score	Match	Length	DB	ID	Description
1	1881.5	46.9	475	11	QBVIEW			QBview0 mus musculus

2	1871.5	46.7	475	11	QBVIEW			QBview1 mus musculus
3	786.5	19.6	277	11	QBVIEW			QBview2 mus musculus
4	636	15.9	419	11	QBVIEW			QBview3 mus musculus
5	440.5	11.0	950	5	QBVIEW			QBview4 drosophila
6	400.5	10.0	429	5	QBVIEW			QBview5 caenorhabdi
7	400.5	10.0	435	5	QBVIEW			QBview6 caenorhabdi
8	371	9.3	571	10	QBVIEW			QBview7 arabisdopsis
9	366	9.1	603	10	QBVIEW			QBview8 arabisdopsis
10	357.5	8.9	532	10	QBVIEW			QBview9 arabisdopsis
11	357	8.9	536	10	QBVIEW			QBview10 arabisdopsis
12	357	8.9	543	10	QBVIEW			QBview11 arabisdopsis
13	346	8.6	251	4	QBVIEW			QBview12 arabisdopsis
14	266.5	6.6	601	10	QBVIEW			QBview13 arabisdopsis
15	253.5	6.3	356	10	QBVIEW			QBview14 arabisdopsis
16	248.5	6.2	658	10	QBVIEW			QBview15 arabisdopsis
17	246.5	6.2	896	10	QBVIEW			QBview16 arabisdopsis
18	228	5.7	526	10	QBVIEW			QBview17 arabisdopsis
19	228	5.7	526	10	QBVIEW			QBview18 arabisdopsis
20	228	5.7	695	10	QBVIEW			QBview19 arabisdopsis
21	228	5.7	695	10	QBVIEW			QBview20 arabisdopsis
22	226.5	5.7	609	10	QBVIEW			QBview21 arabisdopsis
23	226.5	5.7	609	10	QBVIEW			QBview22 arabisdopsis
24	186	4.6	243	5	QBVIEW			QBview23 arabisdopsis
25	183.5	4.6	593	10	QBVIEW			QBview24 arabisdopsis
26	168.5	4.2	553	10	QBVIEW			QBview25 arabisdopsis
27	150.5	3.7	1175	11	QBVIEW			QBview26 oryza sativ
28	146.5	3.6	1175	11	QBVIEW			QBview27 mus musculus
29	146	3.6	1737	4	QBVIEW			QBview28 mus musculus
30	134.5	3.4	4998	11	QBVIEW			QBview29 homo sapien
31	131.5	3.3	1428	5	QBVIEW			QBview30 mus musculus
32	131.5	3.3	4123	4	QBVIEW			QBview31 mus musculus
33	129.5	3.2	1145	11	QBVIEW			QBview32 mus musculus
34	129.5	3.2	1450	11	QBVIEW			QBview33 mus musculus
35	126.5	3.2	1840	4	QBVIEW			QBview34 mus musculus
36	125.5	3.1	1182	4	QBVIEW			QBview35 mus musculus
37	123.5	3.0	1235	4	QBVIEW			QBview36 mus musculus
38	123.5	3.0	4957	4	QBVIEW			QBview37 mus musculus
39	123.5	3.0	5262	4	QBVIEW			QBview38 mus musculus
40	123	3.0	1190	4	QBVIEW			QBview39 mus musculus
41	122.5	3.1	1130	4	QBVIEW			QBview40 mus musculus
42	121	3.0	955	4	QBVIEW			QBview41 mus musculus
43	120.5	3.0	944	11	QBVIEW			QBview42 mus musculus
44	120.5	3.0	946	11	QBVIEW			QBview43 mus musculus
45	120.5	3.0	5146	6	QBVIEW			QBview44 mus musculus

ALIGNMENTS

RESULT 1	QBVIEW	PRELIMINARY;	PRT;	475 AA.
ID	QBVIEW0			
AC	QBVIEW0			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Tesmin (Hypothetical 50.6 kDa protein)			
GN	MTL5 OR TESM.			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/C;			
RA	Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;			
RT	Structure of the tesmin gene encoding a testis-specific persistent			
RT	protein; a possible multifunctional protein with dynamic changes of			
RT	localization throughout spermatogenesis.*;			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			

DR	EMBL; AB057423; BAB64935.1; -.
DR	EMBL; BC024377; AAH24377.1; -.
DR	MGI; MGI:1340029; Mtl5.
DR	InterPro; IPR005172; CXc.
DR	Pfam; PF03638; CXc; 1.
KW	Hypothetical protein.
SQ	SEQUENCE 475 AA; 50615 MW; 4373ED1188DCF9F CRC64;
Alignment Scores:	
Pred. No.:	4.19e-172
Score:	1881.50
Percent Similarity:	96.47%
Best Local Similarity:	95.92%
Query Match:	46.94%
DB:	11
Gaps:	1
US-09-743-237-1 (1-2241) x Q8VIE0 (1-475)	
QY	453 CTACTGCCAGGTTCCCAAGCAGCAACGAAGAAAAAAGT----- 494
DB	LeuLeuProGlyAlaArgGlyProAlaLeuLeuProLeuSerAlaGlyValArgValIle 127
QY	495 ---GTTCAATCAAAGAAGCAGGTGGTAGTGTCGCAGCGCGCACCTCGAAGACGCAGCT 551
DB	ProValGluIleLysGluAlaGlySerValProGlyGlySerProGluAspAlaAla 147
QY	552 TTCAGGCCCTCTGGCTCAGGAATCTGTTCGAAGTCCCATCATCCCCAGGAGGCAG 611
DB	PheGlnAlaProLeuAlaGlnGluSerCysLysPheProSerGlnGluAlaGlu 167
QY	612 GAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCCTGTGTATTGTCAGCTGAAA 671
DB	GluAlaSerSerCysProArgLysLysAspSerSerPrometValIleCysGlnLeuLys 187
QY	672 GGAGGCCCCAGATGCTCTCATAGACAACCTGTGGCCGAGGAGCTCAAGCGCTCCAT 731
DB	GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis 207
QY	732 CTCGTTCTCAGTAGCATGACCAGAGAGTTCCCTCAGTCAGAGCTCCCTAAGCCAATG 791
DB	LeuLeuProGlnTyrraspaspgInSerSerPheProGlnSerGluLeuProLysPromet 227
QY	792 ACACTTTAGTGGGAAGACTCTGCCAGTACCAGCGAAGTTAAATCTCATCACAGGTT 851
DB	ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal 247
QY	852 GATAATGGAGCTCTCCCATCAGCTGTCAATGGGGCTGCCCTTCGCTGACCTGCTCTG 911
DB	AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu 267
QY	912 CAAGGGCCACCCAAAATAACTCTGTCTGGGTACTGTGACTGCTTCTCCACGGGGACTTC 971
DB	GlnGlyProProLysIleThrLeuSerGlyTyrrCysaspCysPheSerSerGlyAspPhe 287
QY	972 TGCAACAGCTGCAGCTGCACAACCTCGCCCATGAGCTCGAGCGCTTCAAGGCCATAAG 1031
DB	CysAsnSerCysSerCysAsnAsnLeuArgHisGluLeuGluArgPheLysAlaIleLys 307
QY	1032 CGGTGCTTGATAGAAATCTCGAGCTTCCACCAAAAATGGGGAAGGCCGCTCGGGA 1091
DB	AlaCysLeuAspArgAsnProGluAlaPheGlnProLysMetGlyLysGlyArgLeuGly 327
QY	1092 GCTGCTAAACTCGACACAGCAAGGTGCAACTGTAAAGCGCTCAGCTGCCTCAAGAAC 1151
DB	AlaAlaLysLeuArgHisSerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsn 347
QY	1152 TACTGTGAGTGCATGAGGCCAAAATCATCTGTTCTTCCATTTGCCAAATGCATGCTTGC 1211
DB	TyrCysGluCysTyrGluAlaLysIleMetCysSerSerileCysLysCysIleAlaCys 367
QY	1212 AAAAATCATGAAGAAAGTCCAGAACGAAAAATGCTGATGAGCACACCCCACTACATGGAG 1271
DB	LvsAsnTyrgluGluSerProGluArqLysMetLeuMetSerThrProHisTyrrMetGlu 387

QY	672	GGAGGCGCCAGATGCTTCATACACAACTCTGGCGGAGGAGCTCAAGCGCTCCAT	731
Db	188	GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis	207
QY	732	CTGCTTCCTCAGTACGATCAGCAGCAGTTCCTCAGTCAGAGCTCCCTAAGCAATG	791
Db	208	LeuLeuProGlnTyrAspAspGlnSerPheProGlnSerGluLeuProLysProMet	227
QY	792	ACAACCTTTAGTGGGAGAGCTTCTGCGAGTACCAGGAGAGTAAATCTCATCACAGGTT	851
Db	228	ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal	247
QY	852	GATATGGAGCTCTCCATCAGCTGTCATGGGGTGGCTTTCCTCTGGAGCTGCTCTG	911
Db	248	AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu	267
QY	912	CAAGGCCACCAAACTACTCTGCTGGTACTGTGACTGCTTCTCCAGCGGGGACTTC	971
Db	268	GlnGlyProProLysIleThrLeuSerGlyTyrCysAspCysPheSerSerGlyAspPhe	287
QY	972	TGCAACAGCTGCGAGCTGCAACCACTGCGCCATGAGCTCGAGCTTCAAGGCCATAAAG	1031
Db	288	CysAsnSerCysSerCysAsnAsnLeuArgHisGluLeuGluArgPheLysAlaLys	307
QY	1032	CGGTGCTTTGATAGAAATCTCTGAAGCTTTCCAAACCAAAATGGGAAAGCGCTGGGA	1091
Db	308	AlaCysLeuAspArgAsnProGluAlaPheGlnProLysMetGlyLysGlyArgLeuGly	327
QY	1092	GCTGTAACTTCGACACAGCAAGGTCACACTGTAAAGCTCAGGCTGCTGAAGAC	1151
Db	328	AlaAlaLysLeuArgHisSerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsn	347
QY	1152	TACTGTGAGTGTATGAGGCCAAATCATGTGCTTCTTCCATTTGCAAAATGCAATTTGCTTC	1211
Db	348	TyrCysGluCysTyrGluAlaLysIleMetCysSerSerIleCysLysCysIleAlaCys	367
QY	1212	AAAACTATGAAGAAGTCCAGAACAAATGCTGATGAGCAGCACCCCACTACATGGAG	1271
Db	368	LysAsnTyrGluGluSerProGluArgLysMetLeuMetSerThrProHisTyrMetGlu	387
QY	1272	CCTGGGAGCTTTGAGCAGCAGCATTTATTTGTCGCCAGCAAGTCTCAGGACCTCCAAA	1331
Db	388	ProGlyAspPheLysSerHisTyrLeuSerProAlaLysPheSerGlyProLys	407
QY	1332	CTGAGAAAATAGCAGGCTTCTCTGTATCTCTCTGGGAAGTAGTGAGGCGCCACATGT	1391
Db	408	LeuArgLysAsnArgGlnAlaPheSerCysIleSerTrpGluValGluAlaThrCys	427
QY	1392	GCCTGCTGCTGGCCAGGCTGAGGAAGCAGAGCAGGAGCAGTGTTCCTCCCAAGCTGGCT	1451
Db	428	AlaCysLeuLeuAlaHisGlyGluGluAlaGluHisGluHisCysSerProSerLeuAla	447
QY	1452	GAGCAGATCATCTCGAGGAGTTTGAAGTGCCTGCTGCGCATTTCTCCATCTCCAGTTC	1511
Db	448	GluGlnMetLeuLeuGluGluPheGlyArgCysLeuSerGlnIleLeuHisIleGluPhe	467
QY	1512	AAGTCCCAAGGGCTGAAATTTGAG	1535
Db	468	LysSerLysGlyLeuLysIleGlu	475
RESULT 3			
Q9D571			
ID	Q9D571	PRELIMINARY;	PRT; 277 AA.
AC	Q9D571		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DE	Adult male testis cDNA, RIKEN full-length enriched library,		
DE	clone:493050902, full insert sequence.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		

RN	[1]	SEQUENCE FROM N.A.	
RP	STRAIN=C57BL/6J; TISSUE=Testis;		
RC	MEDLINE=21085660; PubMed=11217851;		
RX	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,		
RA	Hayashizaki Y.		
RT	"Functional annotation of a full-length mouse cDNA collection."		
RL	Nature 409:685-690(2001).		
DR	EMBL; AK015732; BAB29949.1; -.		
SQ	SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;		
Alignment Scores:			
Pred. NO.:	1.18e-66	Length:	277
Score:	786.50	Matches:	155
Percent Similarity:	92.35%	Conservative:	2
Best Local Similarity:	91.18%	Mismatches:	6
Query Match:	19.62%	Indels:	7
DB:	11	Gaps:	1
US-09-743-237-1 (1-2241) x Q9D571 (1-277)			
QY	453	CTATGCCAGGTTCCCAAGCGCAGCAAGCAAGCAAGCAAGCAAGT-----	494
Db	108	LeuLeuProGlyAlaArgGlyProAlaLeuLeuProLeuSerAlaGlyValArgValIle	127
QY	495	---GTTGAATCAAGAGCAGCGTGTGTGTCCAGCGCGCAGCGCTCAAGCGCAGCT	551
Db	128	ProValGluIleLysGluAlaGlyGlySerValProGlyGlySerProGluAspAlaLa	147
QY	552	TTCCAGGCGCCCTGTGCTCAGCAATCCTGTTCAGTTCCTCATCCAGGAGCAGAG	611
Db	148	PheGlnAlaProLeuAlaGlnGluSerCysLysPheProSerSerGlnGluAlaGlu	167
QY	612	GAGGCTCCAGCTGCGCTCGGAAGAAGACTCCAGCCCCATGTTGTTGTCAGCTGAAA	671
Db	168	GluAlaSerSerCysProArgLysLysAspSerSerProMetValIleCysGlnLeuLys	187
QY	672	GGAGGCGCCAGATGCTCTGCATAGACAACACTGTGGCGCGGAGGCTCAAGCGCTCAT	731
Db	188	GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis	207
QY	732	CTGCTTCCTCAGTACGATGACGAGCAGCTTTCCCTCAGTCAGAGCTCCCTAAGCAATG	791
Db	208	LeuLeuProGlnTyrAspAspGlnSerSerPheProGlnSerGluLeuProLysProMet	227
QY	792	ACAACCTTTAGTGGGAGAGCTTCTGCGAGTACCAGCGCAAGTTAAATCTCATCACAGGTT	851
Db	228	ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal	247
QY	852	GATATGGAGCTCTCCATCAGCTGTCATGGGGTGGCTTTCCTCTCGAGCTGCTCTG	911
Db	248	AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu	267
QY	912	CAAGGCCACCCCAAAATAACTCTGCTGGG	941
Db	268	GlnGlyProProLysIleThrLeuSerGly	277

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., McLeod M.P., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidon-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR FlyBase: A003818; AAF58365.1; -.
 DR FlyBase: FBgn0033846; CG6061.
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 SQ SEQUENCE 950 AA; 10021 MW; 627C3EA6B4A0A30 CRC64;

Alignment Scores:

Pred. No.: 3,68e-33 Length: 950
 Score: 440.50 Matches: 138
 Percent Similarity: 45.87% Conservative: 62
 Best Local Similarity: 31.65% Mismatches: 141
 Query Match: 10.99% Indels: 95
 DB: 5 Gaps: 14

US-09-743-237-1 (1-2241) x Q9V608 (1-950)

QY 257 CCCCTCTGAAGTAAAGGACAGTCTTTAAACCTATGTAGAGTTTGTGATGAATCTGCT 316
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 572 ProValGlnLysLeuIleAsnArgThrAlaAnValGlnArgValSerSerThr 591
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 317 TCTCAACCATATGCTAAGCTATATGCAATCTCTTGAATGCTATATTAATCTAGGAGA 376
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 592 SerPro-Ser-----SerAsnSerThrLysLysIleTyrAsnTyrValGI 606
 QY 377 ACCTCTGATTCTCTGCTCTATCTATCTGCTGCTAGTGTACAGGGGGAATCATTTTG 436
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 606 nPro-----ThrGlySerAsnGlyAlaLysTyrMetIleCysAsnSe 620
 QY 497 TCAAAATCAAGAGCAGGTGGTGTGCCA-----GGCGGCGCCCTGAAGA 544
 DB : : : : : : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 620 rGlyValProGlnSerSerThrSerAlaMetArgGlyThrGlyTyrValGlnAs 640
 QY 545 CCAGCTTTCCAGGCCCTCTGGCTCAGGAATCTGTTGCAAGTCTCCCATCATCCAGGA 604
 DB : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 640 n-LysThrArgPro-----ProProIleSerProGlnGlnHisArgPheL 656
 QY 605 GCGAGGAGGCTCCAGCTGCTGCTCGGAGAAAGACTCCAGCCCATGGTATTGTCA 664
 DB : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 656 yGlnMetGlyPro-GlnGlnGlnSerLys-----HisGlnGlnGlnGln 670
 QY 665 GCTGAAAGGCGGCCAGATGCTCTGCGATAGACAACTGTGGCGGAGGAGCTCAAAGC 724
 DB : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 671 AlaGlnAlaLysGln-----ArgIleArgGlnGlnGlnGlnGlnSer 687
 QY 725 GCTCCATCTGCTTCTCAGTACGATGACACAGCAGAGTTTCCCTCAGTCAGAGCTCCCTAA 784

DB 688 ThrProIleLysValGluProLysLeuProThrLeuProGlyValLysAla-AsnVa 707
 QY 795 GCCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGCAAGTAAATCTCATCAC 844
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 707 IProAlaLysProLeuPheGluValLeuLysPro--ProAla----- 720
 QY 845 ACAGGTTGATAATGAGGCTCTCCCATCAGCTCAATGGGCTGCTTCCCTCTGGACC 904
 DB : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 721 -----ThrAlaAlaAlaAlaGlyAlaValAspProLeuGlyG1 733
 QY 905 TGCTCTGCAAGGGCCACCCAAATAACTGTCTGGG-----TACTG 946
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 733 yMetThrSerArgLysHisCysAsnCysSerLysSerGlnCysLeuLysLeuTyrCy 753
 QY 947 TGACTGCTTCCACGGGGGACTTCTGCAACAGCTGCAGCTGC-----NACAA 994
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 753 sAspCysPheAlaAsnGlyGluPheCysGlnAspCysThrCysLysAspCysPheAsnAs 773
 QY 995 CCTGGCCCATGAGCTCGAGCGCTTCAAGCCATAAGCGCTGTCTGTATAGAAATCCTGA 1054
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 773 nLeuAspTyrGluValGluArgGluArgAlaIleArgSerCysLeuAspArgAsnProSe 793
 QY 1055 AGCTTTCCAAACCAAAATGGGAAAGCGCTCTGGAGCTGCTAAACTTCGACACAGCAA 1114
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 793 rAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetArgLeu---HisAsnLy 812
 QY 1115 AGGTGCAACTGTAAAGCGCTCAGGCTGCTGAAGAACTACTGTGAGTGTATGATGAGGCCNA 1174
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 812 sGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLy 832
 QY 1175 AATCATGTGTTCTTCCATTTGCAATGCTTCTGCAAAACTATGAAGAAAGTCCAGA 1234
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 832 sileProCysSerSerIleCysLysCysValGlyCysArgAsnMetGluAspArgProAs 852
 QY 1235 AGCAAAATGTGTATGATGACACACCCACTACATGAGGCTGTGGGCTTTTGTAGAGAGCCA 1294
 DB : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 852 p-----ValAspMetAspSerLeuAs 859
 QY 1295 TTATTGTTCCCGCAAGTTCTCAGGACCTCCAAACTGAGAAAAATAGACAG----- 1349
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 859 pGlyLeuMetGlyValGluGlyGlnLysLysAspLys---AlaLysAsnLysGlnLeuAs 878
 QY 1350 -----GCCTTCTCTGTATCTCTGGAAGTGTAGGAGGCGCACATGTGCTGCT 1399
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 878 nClusAsnArgAlaAsnIleTyrPheThrAspValIleGluAlaThrIleMetCysMe 898
 QY 1400 GCTGCGCCAGGTGAGGAAGCAGACAGCAGGACACTGTTCCCAAGCTTGGCTGAGCAGAT 1459
 DB : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 898 tileSerArgIleValMetHisGluLysGlnAsnValAlaValGluAspMetGluArgG1 918
 QY 1460 GATCTGAGGAGTTGGNAGGTGCTGTCGAGATTCTC 1499
 DB : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 918 uValMetGluGluMetGlyGluSerLeuThrGlnIleIle 931
 RESULT 6
 O62295 PRELIMINARY; PRT; 429 AA.
 AC O62295;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE JCB 6b protein.
 GN JCB 6 OR JCB 6B.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lightning J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]


```

Best Local Similarity: 31.13%      Mismatches: 157
Query Match:          9.98%      Indels: 78
DB:                   5          Gaps: 18

US-09-743-237-1 (1-2241) x Q950D7 (1-435)

QY 419 CAGGGGAAATCATTTTGTGTGACACTCCGATGAACACTACTGCCAGGTTCCTCAAGGCGAGCAA 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 GlnGlyGluIleVal-TyrGlnAspAspAspTyrTyrAspGluSerGluIleTyrAs 22

QY 479 GCAAGCAACAAAGTGTGAATCAACAAAGCAGGTGTGTAGTGTGCCAGGCGGCGACCC 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 22 pAsnTyrGluGluGlyAlaGluPheIleGluValAsnGluGlnLeuValProHisAsnPr 42

QY 539 TGAAGACGCGAGCTTTCCAGGCGCCCTCTGGCTCAGGAATCCTGTGGCAAGTTCCTCATCATC 598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 42 O-AsnLeuGlnAlaGlnGln-----AsnArgProGlyThrSerSerMetIleG 58

QY 599 CCAGGAGGAGGAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCGCCCTGGTGTGAT 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 InGlnHisAsnArgSerMetGluValAsnGlnGlnGlyLeuValLysAspGluProIle-Asp 77

QY 659 TTGTACAGTCAAGG-----AGGCCCCAGATGCTCTCATAGACAACTGTGCCGC 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 ThrSerSerHisArgValTyrValProProProArgProValGlnArgLysLeuPheGln 97

QY 710 GAGGAGCTCAAGGCTCCATCTGCTCTCAGTAC-----GATGACCA 754
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 ProGly---ProSerThrProGlySerSer-GlnTyrThrValArgAsnLeuSerAsnLe 116

QY 755 GAGCAGTTCCTCCAG-----TCAGAGCTCCCTAAG-----CC 787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116 uSerGlySerProSerMetTyrAspArgGlnProAlaSerLeuProArgThrValGlnPr 136

QY 788 AATGACAACTTAGTGGGAAGACTTCTGCCAGTACCAGCGAAGTAAATCTCATCACACA 847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 oMetGlyLeuGluMetGlyAsnSerGluGlnArgLysValTyrIleAspMetLysAspH 156

QY 848 GGTGATAATGAGCTCTCCCATCAGCTGCTCAATGGGCTGCTTCCCTCTGGAGCTGC 907
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 sValSerHisIleArgLeuLysThrLysLysValPheAla-----ProG 172

QY 908 TCTGCAAGGCGCACCAAAATAACT-----CTGCTGGGTACTGTGAGCTGT 955
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 yGlnArgLysProCysAsnCysThrLysSerGlnCysLeuLysLeuTyrCysAspCysPh 192

QY 956 CTCCAGCGGGACTTCTGCAACAGCTGCAGCTGC-----AACAACTGCGCCA 1003
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 eAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHisAsnAsnIleGluTy 212

QY 1004 TGAGCTCGAGCGCTTCAAGCCATAAAGCGTGTCTGTATAGAAATCTCGAAGCTTTCCA 1063
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 rAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsnProAsnAlaPheLy 232

QY 1064 ACAAAAATGGGAAGCGGCTCTGGGAGCTGCTAAACTTGA-----CACAGCAAAAGG 1117
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 sProLysIleGlyIleAlaArgGlyGlyIleThrAspIleGluArgLeuHisGlnLysG 252

QY 1118 GTGCAACTTAAGCGCTGAGCTGCTGAAGAACTACTGAGTGTATGAGCGCAAAAT 1177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 yCysHisCysLysLysSerGlyCysLeuLysAsnTyrCysGluCysTyThrGluAlaLysVa 272

QY 1178 CATGTTCTTCCATTGCAATGCAATGCTTGTGCAAAAC-----1217
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 272 lProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGluThrTyrArgMetTh 292

QY 1218 ----TATGAAGAAAGT-----CCAGAACGAAAAATGCTGATGAGC-----1253
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 rArgTyrLysAsnSerGlyAlaValSerAsnThrAsnAlaLeuMetSerLeuThrAs 312

QY 1254 -----ACACCCCATACATGAGCGCTGGGACTTTGAGCAGCAGCATTA 1297
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 nAlaSerSerThrAlaThrProAsp---SerGlyProGlySerValValThrAspGluH 331

```

```

1298 T-----TTGTCCCGCAGCAAGTTCTCAGGACCTCCAAAACCTGAGAAAAAATAGGCA 1348
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
331 sGlyAspAspTyrGluAspMetLeuLeuSerHisLysProLysValGluMetAspProAr 351

1349 GGCCTTCTCTCTATCTCTCTGG-----GAAGTAGTGGAGGCCACATGTGC 1393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 gArgPhe-----ProTyrTyrTyrMetThrAspGluValValGluAlaAlaThrMe 368

1394 CTGCGCTTGGCCCGAGGCTGAGGAACA-----GAGCAGGAGCACTG 1435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
368 tCysMetValAlaGlnAlaGluAlaLeuAsnTyrGluLysValGlnThrGluAspG 388

1436 TTCCCCAAGCTTGGCTGAGCAGATGATCTGGAGGAGTTTGGAGGTGCTGTCGCAGAT 1495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
388 uLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGluGlnMe 408

1496 TCTC 1499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
408 tile 409

RESULT 8
Q9SL70 PRELIMINARY; PRT; 571 AA.
ID. Q9SL70; AC Q9SL70;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE At2g20110 protein.
GN At2G20110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.-J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006081; AAD24386.1; -
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXG; 2.
SQ SEQUENCE 571 AA; 61312 MW; 36701C5FA62B3B89 CRC64;

Alignment Scores:
Pred. No.: 1,58e-26 Length: 571
Score: 371.00 Matches: 92
Percent Similarity: 45.85% Conservative: 46
Best Local Similarity: 30.56% Mismatches: 101
Query Match: 9,26% Indels: 62
DB: 10 Gaps: 11

US-09-743-237-1 (1-2241) x Q9SL70 (1-571)

QY 603 GAGCAGAGGAGGCGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGTGTTGT 662
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 GluGlyGluGlyAspLysPheProLysThrAspGlu-----ValThrGln 19

```



```

Db      142 LeuLysLeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCysAsnCysVal 161
QY      990 -----AACACCTGGCCCATGAGCTCGAGCGCTTCAAGCCATAAAGCGTCTCTT 1040
Db      162 AsnCysPheAsnAsnValAspAsnGluProAlaArgArgGluAlaValGluAlaThrLeu 181
QY      1041 GATAGAAATCCTGAAGCTTCCAAACCAAAATGGGAAA-----GCCCGT--- 1085
Db      182 GluArgAsnProPheAlaPheArgProLysIleAlaSerSerProHisGlyGlyArgAsp 201
QY      1086 -----CTGGAGCTGCTAACTT-----CGACACGACAAAGGTCAACTGT 1127
Db      202 LysArgGluAspIleGlyGluValValLeuLeuGlyLysHisAsnLysGlyCysHisCys 221
QY      1128 AAGCGCTCAGGCTCGTGAAGAACTACTGTAGTGTATGAGCCCAAAATCATGTCTTCT 1187
Db      222 LysLysSerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIleLeuCysSer 241
QY      1188 TCATTTCGAAATGCAATTCGTCGAAAACTATGAAGAAAGTCCAGAACGAAAAATGCTG 1247
Db      242 GluAsnCysLysCysLeuAspCysLysAsnPheGluGlySerGluGluArgGlnAlaLeu 261
QY      1248 ATGAGCACACCCAC-----TACATGGAG----- 1271
Db      262 PheHisGlyGluHisSerAsnHisMetAlaTyrLeuGlnGlnAlaAlaAsnAlaIle 281
QY      1272 CCTGGGGACTTTCAGAGCAGCCATTATTG---TCCCGCCAGCTTCTCAGGACCTCCA 1328
Db      282 ThrGlyAlaValGlySerGlyPheAlaProSerProAla-----Pro 296
QY      1329 AAACCTGAGAAAAAATAGGCAG 1349
Db      297 LysArgArgLysGlyGlnGlu 303

RESULT 10
Q8GUZ0
ID      Q8GUZ0      PRELIMINARY;      PRT;      532 AA.
AC      Q8GUZ0;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Hypothetical protein (Fragment).
GN      T2617.9.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Xiao Y., Ishmael N., Kumar N., Redman J., Smith S., Riedmuller S.,
RA      Uterback T., Whitelaw C.A., Fraser C.M., Town C.D.;
RT      "Cloning and sequencing of full-length cDNAs for hypothetical genes
RL      from chromosome 2 of Arabidopsis thaliana."
RW      Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AV163775; AACN5199.1;
KW      Hypothetical protein.
FT      NON_TER
SQ      SEQUENCE      532 AA;      57033 MW;      78CF4E441175958D CRC64;

Alignment Scores:
Pred. No.:      3,1e-25      Length:      532
Score:      357.50      Matches:      102
Percent Similarity:      40.62%      Conservative:      54
Best Local Similarity:      26.56%      Mismatches:      114
Query Match:      8.92%      Indels:      114
DB:      10      Gaps:      12

US-09-743-237-1 (1-2241) x Q8GUZ0 (1-532)
QY      756 AGCAGTTTCCTCAGTCAGAGCTCCCTTAAGCCAAATGACAACTTTAGTGGGAAGACTTCTG 815
      |||||      : : : : : |||||      : : : : :

```

```

Db      11 SerSerMetAlaAlaAlaSerIleProSerProIleValThrValThrArgProIleIle 30
QY      816 -----CCAGTACCAGCAAGTTAAAT 836
Db      31 ThrSerGlnAlaProProThrValAlaThrProIleProProGlnSerGlnGly 50
QY      837 CTATCACACAGCGTTGAT---AATGGAGCTCTCCATCAGCTGTCAATGGCGTGCCTTT 893
Db      51 IleIleLeuHisValProIleArgHisProArgProGluSerProAsnSerMetProArg 70
QY      894 CCCTCTGGGACCTCTCTCGAAGGGCCACCCAAA----- 926
Db      71 ProAlaGlyLulThrArgAspGlyThrProGlnLysLysGlnCysAsnCysLysHis 90
QY      927 ---ATAACTCTGCTGGTACTCTGACTGCTTCCAGCGGGGACTTCTCAACAGCTGC 983
Db      91 SerArgCysLeuLysLeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCys 110
QY      984 AGCTGC-----AACAACTCGCGCATGAGCTCGAGCGCTTCAAGCCATAAAG 1031
Db      111 AsnCysValAsnCysPheAsnAsnValGluAsnGluProAlaArgGlnAlaValGlu 130
QY      1032 GCCTGCTTCATGAATACTCTGAAGCTTCCCAACCAAAATGGGAAAGCCCTCTGGGA 1091
Db      131 SerThrLeuGluArgAsnProAsnAlaPheArgProLysIleAlaAlaSerProHisGly 150
QY      1092 GCTGCTAAA-----CTTCACACACAAAGG 1118
Db      151 GlyArgAspAsnArgGluGluValGlyAspValValMetLeuAlaArgHisAsnLysGly 170
QY      1119 TGCAACTGTAAAGCGCTCAGCTCGCTGAAGAACTACTGTGAGTGTATGAGGCAAAATC 1178
Db      171 CysHisCysLysLysSerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIle 190
QY      1179 ATGTGTTCTTCATTTGCAAAATGCATGCTTGCATAAACTATGAAGAAATCCAGAACGA 1238
Db      191 LeuCysSerGluAsnCys***CysLeuAspCysLysAsnPheGluGlySerGluValArg 210
QY      1239 AAAATGCTGATGACGACACCCAC-----TACATGGAG----- 1271
Db      211 GlnSerLeuPheHisGlyGluHisSerHisAsnLeuAlaTyrLeuGlnHisAlaAsnAla 230
QY      1272 -----CCTGGGAGCTTTGAGAGCGCCATTATTGTGCCCGCCAGCTTCTCAGGACCT 1325
Db      231 AlaIleThrGlyAlaIleGlySerSerGlyPheAlaSer-----AlaProPro 246
QY      1326 CCAAACTGAGAAAAAATAGGAGCGCTTC----- 1355
Db      247 ProLysArgArgLysGlyGlnGluIlePhePheAsnGlnGlyThrLysAspSerSerThr 266
QY      1355 ----- 1355
Db      267 HisArgLeuGlyGlnAlaAsnAsnGlyArgThrThrSerSerGlnThrGlySerArgAla 286
QY      1356 -----TCCTGTATCTCTCTGGAA 1373
Db      287 GlyGlyAsnAlaSerLeuGlyProSerLysSerLeuLeuAlaAsnIleLysProMet 306
QY      1374 GTAGTGGAGGCCACATGTGCTGCTGCCCGAGGCTGAGGAGCAGAGCAGGACCAC 1433
Db      307 AspValLysAlaLeuCysSerValValAlaValAlaGlyGluAlaAlaLysThrLeu 326
QY      1434 TGTTCGCCAAGCTTGGCTGAGCAG-----ATGATCTCGGAGGACTTT 1475
Db      327 ThrGluLysArgLeuAlaAsnGlnLysGluThrSerValAlaSerSerValGlnAspGln 346
QY      1476 GGAAGGTGCTGTGCGCAGATTCTCCACATCAGTTCCAAGTCCCAAG-----GGGCTGAAA 1529
Db      347 Gly-----HisValAsnAsnLysAlaGluLysSerGlyLeuGlu 359
QY      1530 ATTGAGTAGCTGCAAGCTGGTAAAGGGGAATGCCTCTGGCAGGCTCAGCCCTGGGAAT 1589
Db      360 Asp-AsnAsnAspGlySerLysGlyArgSerLeuSerProGluThrLeuAlaLeuMe 379
      |||||      : : : : : |||||      : : : : :

```

```
QY 1590 CPGCACCAG 1599
Db 379 tCysAspGlu 382

RESULT 11
Q8GUZ2 PRELIMINARY; PRT; 536 AA.
AC Q8GUZ2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN T2G17.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao Y., Ishmael N., Kumar N., Redman J., Smith S., Riedmuller S.,
RA Utterback T., Whitelaw C.A., Fraser C.M., Town C.D.;
RT "Cloning and sequencing of full-length cDNAs for hypothetical genes
RT from chromosome 2 of Arabidopsis thaliana.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY163773; AAN85197.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 536 AA; 57522 MW; 3A841C7222010E08 CRC64;

Alignment Scores:
Pred. No.: 3,478-25 Length: 536
Score: 357.00 Matches: 80
Percent Similarity: 47.60% Conservative: 39
Best Local Similarity: 32.00% Mismatches: 77
Query Match: 8.91% Indels: 54
DB: 10 Gaps: 8

US-09-743-237-1 (1-2241) x Q8GUZ2 (1-536)
QY 756 AGCAGTTCCTCAGTCAGCTCCCTAAGCCAAATGACAACCTTTAGTGGGAGACTTCTG 815
Db 11 SerSerMetAlaAlaAlaSerIleProSerProIleValThrValThrArgProIleIle 30
QY 816 -----CCAGTACCAGCGAAGTTAAAT 836
Db 31 ThrSerGlnAlaProProThrValAlaThrProIleProProGlnSerGlnGly 50
QY 837 CTCATCACACAGTTGAT---AATGGAGCTCTCCCATCAGCTGTCAATGGGCTCCCTTT 893
Db 51 IleIleLeuHisValProIleArgHisProArgProGluSerProAsnSerMetProArg 70
QY 894 CCCTCTGGACCTGCTCTCAAGGGCCACCCAAA----- 926
Db 71 ProAlaGlyGluThrArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHis 90
QY 927 ---ATAACTCTGCTGGTACTGTGCTTCTCCAGCGGGGACTCTGCAACAGCTGC 983
Db 91 SerArgCysLeuLysLeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCys 110
QY 984 AGCTGC-----AACAACTGGCGCTGAGCTCGAGCGCTTCAAAAGCCATAAAG 1031
Db 111 AsnCysValAsnCysPheAsnValGluAsnGluProAlaArgArgGlnAlaValGlu 130
QY 1032 CGGTCTCTGATAGAAATCTCGAAGCTTTCACCAACCAAAANTGGGAAAGCGCTCGGA 1091
Db 131 SerThrLeuGluArgAsnProAsnAlaPheArgProLysIleAlaAlaSerProHisGly 150
QY 1092 GCTGCTAAA-----CTTCGACACAGCAAGGG 1118
Db 151 GlyArgAspAsnArgGluGluValGlyAspValValMetLeuAlaAlaArgHisAsnLysGly 170

RESULT 12
Q8GUZ1 PRELIMINARY; PRT; 543 AA.
AC Q8GUZ1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN T2G17.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao Y., Ishmael N., Kumar N., Redman J., Smith S., Riedmuller S.,
RA Utterback T., Whitelaw C.A., Fraser C.M., Town C.D.;
RT "Cloning and sequencing of full-length cDNAs for hypothetical genes
RT from chromosome 2 of Arabidopsis thaliana.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY163774; AAN85198.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 543 AA; 58302 MW; AAAA32AD55102F41 CRC64;

Alignment Scores:
Pred. No.: 3,498-25 Length: 543
Score: 357.00 Matches: 80
Percent Similarity: 47.60% Conservative: 39
Best Local Similarity: 32.00% Mismatches: 77
Query Match: 8.91% Indels: 54
DB: 10 Gaps: 8

US-09-743-237-1 (1-2241) x Q8GUZ1 (1-543)
QY 756 AGCAGTTCCTCAGTCAGCTCCCTAAGCCAAATGACAACCTTTAGTGGGAGACTTCTG 815
Db 11 SerSerMetAlaAlaAlaSerIleProSerProIleValThrValThrArgProIleIle 30
QY 816 -----CCAGTACCAGCGAAGTTAAAT 836
Db 31 ThrSerGlnAlaProProThrValAlaThrProIleProProGlnSerGlnGly 50
QY 837 CTCATCACACAGTTGAT---AATGGAGCTCTCCCATCAGCTGTCAATGGGCTCCCTTT 893
Db 51 IleIleLeuHisValProIleArgHisProArgProGluSerProAsnSerMetProArg 70
QY 894 CCCTCTGGACCTGCTCTCAAGGGCCACCCAAA----- 926
Db 71 ProAlaGlyGluThrArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHis 90
QY 927 ---ATAACTCTGCTGGTACTGTGCTTCTCCAGCGGGGACTTCTGCAACAGCTGC 983
```



```

Db      91 SerArgCysLeuLysLeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCys 110
QY      984 ACCTGC-----AACAACTGCCCATGAGCTGAGCGCTTCAAGCCATAAG 1031
Db      111 AsnCysValAsnCysPheAsnAsnValGluAsnGluProAlaArgGlnAlaValGlu 130
QY      1032 GCCTGCTGTATAGAACTCTCAAGCTTTCCAAACCAAAATGGGAAAGCGCTCTGGGA 1091
Db      131 SerThrLeuGluArgAsnProAsnAlaPheArgProLysIleAlaAlaSerProHisGly 150
QY      1092 GCTGCTAAA-----CTTCGACACAGCAAAAGG 1118
Db      151 GlyArgAspAsnArgGluGluValGlyAspValMetLeuAlaArgHisAsnLysGly 170
QY      1119 TCAAACTGTAAGCGCTGAGCTGCTGAGAACTACTGTGAGTGTATGAGCGCCAAATC 1178
Db      171 CysHisCysLysLysSerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIle 190
QY      1179 ATGTGCTTCTTCCATTGCAAAATGCTTGTGCAAAACTATCAAGAAAGTCCAGAAACGA 1238
Db      191 LeuCysSerGluAsnCys**CysLeuAspCysLysAsnPheGluGlySerGluValArg 210
QY      1239 AAAATGCTGATGAGCACACCCAC-----TACATGGAG----- 1271
Db      211 GlnSerLeuPheHisGlyGluHisSerHisAsnLeuAlaTyrLeuGlnHisAlaAsnAla 230
QY      1272 -----CCTGGGGACTTTGAGAGAGCCCATTTGTTCCCGCCAGGTTCTCAGGACCT 1325
Db      231 AlaIleThrGlyAlaIleGlySerGlyPheAlaSer-----AlaProPro 246
QY      1326 CCAAACTGAGAAAATAGGAGGCGCTTC 1355
Db      247 ProLysArgArgLysGlyGlnGluIlePhe 256

RESULT 13
Q8NCC8 PRELIMINARY; PRT; 251 AA.
AC Q8NCC8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE Hypothetical protein FLN90336.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP
RA Isoai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Kotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074817; BAC11228.1; -.
KW Hypothetical protein.
SQ SEQUENCE 251 AA; 26737 MW; 4D37EB2616DD1106 CRC64;

Alignment Scores:
Pred. No.: 3.2e-24 Length: 251
Score: 346.00 Matches: 66
Percent Similarity: 80.65% Conservative: 9
Best Local Similarity: 70.97% Mismatches: 18
Query Match: 8.63% Indels: 0
DB: 4 Gaps: 0

```

```

US-09-743-237-1 (1-2241) x Q8NCC8 (1-251)
QY      495 GTTGAATCAAGACAGGCTGTAGTGTGCCAGCGCGCCCTGAAGACGCGCTTC 554
Db      158 ValGluIleLysGluAlaGlyGlyThrThrSerAsnAsnProGluGluAlaThrLeu 177

```

```

QY      555 CAGGCCCTCTGCTCAGGAATCCTCTTCAAGTTCCCATCATCCAGGAGCAGGAG 614
Db      178 GlnAsnLeuLeuAlaGlnGlnSerCysLysPheProSerSerGlnLeuGluAsp 197
QY      615 GCCTCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTGTATTGTGTCAGCTGAAAGCA 674
Db      198 AlaSerCysCysSerLeuLysLysAspSerAsnProMetValIleCysGlnLeuLysGly 217
QY      675 GCGGCCCATGCTCTGCTGATAGCAACTGTGGCGGAGGAGGAGCTCAAGCGCTCATCTG 734
Db      218 GlyThrGlnMetLeuCysIleAspAsnSerArgThrArgLysLeuAlaLeuHisLeu 237
QY      735 CTTCTCTCAGTACGATGACCAAGCAGCAGTCTTCCCTCAGTCA 773
Db      238 ValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSer 250

RESULT 14
Q9CAV1 PRELIMINARY; PRT; 601 AA.
AC Q9CAV1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical 66.8 kDa protein.
GN T9J14.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Partmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisme N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmlebach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
RA Reichelt J., Schafte M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauró C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monnhaupt G., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Waits A., Utterback T., Fujii C.Y., Shea T.P.,
RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Millischer J., Sellers P., Gill J.E., Feldblum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana.";
RL Nature 408:820-822(2000).
DR EMBL; AC009465; AAG51411.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXG; 2.
KW Hypothetical protein.
SQ SEQUENCE 601 AA; 65786 MW; 10848E970D81E022 CRC64;

Alignment Scores:
Pred. No.: 1.89e-16 Length: 601
Score: 266.50 Matches: 89

```

```

Percent Similarity: 37.77%      Conservative: 33
Best Local Similarity: 27.55%    Mismatches: 96
Query Match: 6.65%             Indels: 105
DB: 10                          Gaps: 16

US-09-743-237-1 (1-2241) x Q9LW71 (1-601)
Qy 495 GTTGAATCAAGAAGCAGGAGTGTGTGCGAGCGCGAGCCCTGAAGAGCGCAGCTTTC 554
    ||||| :|||: |||||
Db 236 ValGluArgGluGlySerCysSerVal-----GlnValAlaAla 249
Qy 555 CAGGCCCTCTGGCTCAGGAATCTGT-----TCAAGTCCCATCATCCAG 602
    ||||| :|||: |||||
Db 250 GlyAlaProAspIleAsnLeuSerCysSerSerLysValAlaAlaIleAspSerThrAla 269
Qy 603 GAGCAGGAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTATTGT 662
    ||||| :|||: |||||
Db 270 GluAlaGluAspLysGluAsp-----LysAspLeuGlnPro----- 281
Qy 663 CAGCTGAAGAGGAGCGCCGATGCTGTGCATAGACAACTGTGCGGAGGAGCTCAAA 722
    ||||| :|||: |||||
Db 282 -----SerGlyLysGlnArgSerValArgArgCysLeuThrPheAspMetGly 298
Qy 723 GCGTCCAT-----CTGCTTCCTCAGTACATGACAGCAGCTTCCCTCAG----- 770
    ||||| :|||: |||||
Db 299 GlySerHisLysArgIleProLeuArgAspSerThrAsnAspLeuProLeuAspSerThr 318
Qy 771 -----TCAGAGCTCCCTAAGCCCAATGACAACCTTTA-----GTGGGA 806
    ||||| :|||: |||||
Db 319 SerIleAsnLysAlaProSerProGlnAsnCysLeuAspThrSerLysGlnAspThrAsp 338
Qy 807 AGACTTCTGCAGTACCAGCGAAGTAAATCTATCATCACAGAGTTGATGGAGCTCTC 866
    ||||| :|||: |||||
Db 339 GluIleLeuProIleProArgThrIleGlyLeuHisLeu-----AsnGlyPheVal 355
Qy 867 CCATCAGCTGTC-----AATGGGCTGCCCTTCCC 896
    ||||| :|||: |||||
Db 356 AsnProSerValSerSerGlyArgLysLysLysLysLysLysAspGlnAlaPhePro 375
Qy 896 ----- 896
Db 376 SerThrThrPheHisTyrAsnIleGluAspGluPheSerThrProValSerThrLysArg 395
Qy 897 -----TCTGGACCTGCTCTGCAAGCGGCCACCAAAATAACTCTCTCTGGG 941
    ||||| :|||: |||||
Db 396 AspLeuValValPheSerAspValLysIleMetGluProGluArgSerValGluGly 415
Qy 942 -----TACTGTGACTGCTTCTCCAGCGGG 965
    ||||| :|||: |||||
Db 416 GluCysPheAspGlnLeuMetAlaMetGluAsnArgTyrCysGluCysPheSerAlaGly 435
Qy 966 GACTTCTGCAACAGC-----TGCAAGTGCACCAACCTG-----CGCCATGAG--- 1007
    ||||| :|||: |||||
Db 436 LeuPheCysGlyGluProCysSerCysGlnAsnCysPheAsnLysProIleHisGluAsp 455
Qy 1008 -----CTCAGCGCTTCAAGCCATGAAGCGGTCTTGTATAGAAATCTCGAAGCTTC 1061
    ||||| :|||: |||||
Db 456 LeuValMetLysSerArgGluValIleLysAla-----ArgAsnProLeuAlaPhe 472
Qy 1062 CAACCAAAATG-----GGAAAGCGCGCTG 1088
    ||||| :|||: |||||
Db 473 AlaProLysValValSerThrSerAspThrValIleAspLeuThrPValGluAsnSerLys 492
Qy 1089 GGAGCTGCTAAACTTCGACACAGCAAGGGTGAACCTGTGAAGCGCTCAGGCTCGCTGAAG 1148
    ||||| :|||: |||||
Db 493 ThrProAlaSerAlaArgHisLysArgGlyCysAsnCysArgLysSerGlyCysSerLys 512
Qy 1149 AACTACTGTAGTCTATGAGCCCAAAATCATGTGTCTTCCATTTCCAAATGATGCT 1208
    ||||| :|||: |||||
Db 513 LysTyrCysGluCysPheMetMetGlyValGlyCysSerSerAsnCysArgCysMetGly 532
Qy 1209 TGCAAAAAC 1217
    ||||| :|||: |||||

```

```

Db 533 CysLysAsn 535
RESULT 15
Q9LW71
ID Q9LW71 PRELIMINARY; PRT; 356 AA.
AC Q9LW71;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genomic DNA, chromosome 3, Pl clone: MSL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RL MEDLINE=20277480; PubMed=10819329;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res.: 7:131-135(2000).
DR EMBL; AB012247; BAB02682.1; -.
DR InterPro; IPR005172; CX.
DR InterPro; IPR000345; CytC_heme_bind.
DR Pfam; PF03638; CX; 2.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
SQ SEQUENCE 356 AA; 39796 MW; 4FE0EA2D9DDBF58 CRC64;

Alignment Scores:
Pred. No.: 2,9e-15 Length: 356
Score: 253.50 Matches: 77
Percent Similarity: 44.58% Conservative: 30
Best Local Similarity: 32.08% Mismatches: 90
Query Match: 6.32% Indels: 43
DB: 10 Gaps: 11

US-09-743-237-1 (1-2241) x Q9LW71 (1-356)
Qy 942 TACTGTGACTGCTTCTCCAGCGGGACTTCTGCAACAGCTGCGCTGCAACACCTGCGC 1001
    ||||| :|||: |||||
Db 79 TyrCysAspCysPheAlaSerGlyValValCysThrAspCysAspCysValAspCysHis 98
Qy 1002 CATGAGCTCGAGCGCTTCAAGCCATAAAGCGGTG-----CTTGATAGAAAT 1049
    ||||| :|||: |||||
Db 99 AsnAsnSerGluLysCysAspAlaArgGluAlaAlaMetValAsnValLeuGlyArgAsn 118
Qy 1050 CCTGAACCTTTCACCAACCAAAATGGGAAAGCGCTCTGGAGCTGTAACTTCGACAC 1109
    ||||| :|||: |||||
Db 119 ProAsnAlaPheSerGluLys---AlaLeuGlySerLeuThrAspAsnGlnCysLysAla 137
Qy 1110 -----AGCAAGGGTGCAACTCTAAGCGCTCAGCGTGC 1142
    ||||| :|||: |||||
Db 138 AlaProAspThrLysProGlyLeuLeuSerArgGlyCysLysCysLysArgThrArgCys 157
Qy 1143 CTGAAGAACACTCTAGTGTCTATGAGCCCAAAATCATGTGTTCTTCATTTGCAATGC 1202
    ||||| :|||: |||||
Db 158 LeuLysLysTyrCysGluCysPheGlnAlaAsnLeuLeuCysSerAspAsnCysLysCys 177
Qy 1203 ATTCTCTTGCAAAATATGAAGAAAT-----CCAGAACGAAAAATGCTGATGAGCACA 1256
    ||||| :|||: |||||
Db 178 IleAsnCysLysAsnValSerGluAlaPheGlnProProAlaPheSerAlaHisAsnSer 197
Qy 1257 CCCCAC---TACATGGAGCGCTTGGGAGAGCCATTTGAGAGAGCCCATTTATTTGTC---CCAGCC 1310
    ||||| :|||: |||||
Db 198 ProGlnValTyrArgArgArgArgArgGluLeuThrGluTrpAsnSerCysProAla 217

```

```
QY 1311 AAG---TTCTCAGGACCT----- 1325
      ||||| |||
Db 218 ProLeuPheSerIleProAspAsnSerIleGlnAsnAlaLeuGlySerProMetSerCys 237
QY 1326 ---CCAAACTG-----AGAAAAATAGGCAGGCTTCTCTGTATCTCTCTGGGAAGTA 1376
      ||||| |||
Db 238 SerProLysLeuProTyrArgLysLysArgSerLeuMetGlyTyrThrSerThrLeuLeu 257
      ||||| |||
QY 1377 GTGGAGGCC-----ACATGTGCTGCTGCTGCCAGGCGTGAAGCA----- 1421
      ||||| ||||| ||| |||
Db 258 ProAspLeuGlyAspLeuCysSerLeuLeuValAlaAlaSerGluSerAlaThrThr 277
QY 1422 -----GAGCAGGAGCACTGTTCCCAAGCTTGGCTGAGCAGATGATCTCGAGGAGTTT 1475
      ||||| ||||| |||||
Db 278 AlaGluAspGlnAsnArgIlePheThrLysProAspAspLysGluAlaIleGluLeuSer 297
QY 1476 GGAAGGTGCTGTCAGATTCTCCACATCGAGTTCAGTCCAGGGGCTGAAATTTGAG 1535
      ||||| ||||| ||||| |||||
Db 298 SerGluSerGluSerArgAsnValGluGluGluIleGlnSerArgGlyArgLeuIleGlu 317
```

Search completed: July 24, 2003, 14:29:59
Job time : 210.573 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:47:53 ; Search time 35,5999 seconds
(without alignments)
5920.628 Million cell updates/sec

Title: .. US-09-743-237-1
Perfect score: 4008
Sequence: 1 latctctgggttgcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp

-MODEL-frame_n2p.model -DEV-xlp
-Q/cn2_1/USPTO.spool/US09743237/runat_24072003_113104_17660/app_query.fasta_1.4750
-DB-SwissProt_41 -Qfmt-fastan -SUFFIX-rsp -MINMATCH=0.1 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 @CGN_1_1_111 @runat_24072003_113104_17660 -NCPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6
-FGAPOP=7 -YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41.1*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	39.9	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
2	1251	31.2	299	1 MTL5_HUMAN	Q9y415 homo sapien
3	144	3.6	1183	1 DRPL_RAT	P54258 rattus norv
4	125.5	3.1	892	1 HIC1_MOUSE	Q9riy5 mus musculus
5	125.5	3.1	1185	1 DRPL_HUMAN	P54259 homo sapien
6	117.5	2.9	309	1 HXA4_CHICK	P17277 gallus gall
7	117.5	2.9	938	1 HDA7_MOUSE	Q8c2b3 mus musculus
8	117.5	2.9	971	1 RECK_HUMAN	Q95980 homo sapien
9	117.5	2.9	1700	1 BAR3_CHITE	Q03376 chironomus
10	115.5	2.9	1509	1 GSRL_HUMAN	Q9nzm4 homo sapien
11	113	2.8	643	1 CD93_RAT	Q9et61 rattus norv
12	112.5	2.8	323	1 AMFR_HUMAN	P26442 homo sapien
13	112.5	2.8	2768	1 THYG_HUMAN	P01266 homo sapien
14	112	2.8	652	1 CD93_HUMAN	Q9npy3 homo sapien
15	111.5	2.8	946	1 IP3L_HUMAN	P27987 homo sapien
16	110.5	2.8	493	1 OC90_HUMAN	Q02509 homo sapien
17	109.5	2.7	478	1 EGR4_RAT	Q00911 rattus norv
18	108.5	2.7	610	1 LEM2_HUMAN	P16581 homo sapien

c 19	108.5	2.7	1427	1 ZFH2_HUMAN	Q9c0a1 homo sapien
c 20	108	2.7	675	1 YMV2_CAEEL	P34504 caenorhabdi
c 21	107.5	2.7	1093	1 SM5B_MOUSE	Q60519 mus musculu
c 22	107	2.6	1142	1 MGCI_HUMAN	O60732 homo sapien
c 23	106.5	2.6	1436	1 WC11_BOVIN	P30205 bos taurus
c 24	106.5	2.7	2805	1 MAPA_HUMAN	P78559 homo sapien
c 25	106	2.6	470	1 EGR2_MOUSE	P08152 mus musculu
c 26	105.5	2.6	611	1 LEM2_CANFA	P33300 canis famli
c 27	105.5	2.6	773	1 MES2_CAEEL	O17514 caenorhabdi
c 28	105.5	2.6	1198	1 HCN4_RAT	Q9fka7 rattus norv
c 29	104.5	2.6	545	1 AIRE_HUMAN	O43381 homo sapien
c 30	104.5	2.6	760	1 P21_XENLA	P16143 xenopus lae
c 31	104.5	2.6	3149	1 TEGU_EBV	P03186 epstein-bar
c 32	104	2.6	1082	1 NCO3_RAT	Q9epu2 rattus norv
c 33	104	2.6	1093	1 AF17_HUMAN	P55198 homo sapien
c 34	104	2.6	1096	1 KPCL_ASPNG	Q00078 aspergillus
c 35	104	2.6	1679	1 FUR2_DROME	P30432 drosophila
c 36	104	2.6	1724	1 PPOV_HUMAN	Q9ukk3 homo sapien
c 37	104	2.6	1964	1 NTC4_MOUSE	P31695 mus musculu
c 38	104	2.6	5376	1 ZAN_MOUSE	O88799 mus musculu
c 39	103	2.6	579	1 SOC6_MOUSE	O8vha2 mus musculu
c 40	103	2.6	1139	1 KPCL_TYRE	Q99014 trichoderma
c 41	103	2.5	1801	1 LMB2_RAT	P15800 rattus norv
c 42	102	2.5	470	1 EGR2_RAT	P51774 rattus norv
c 43	102	2.5	867	1 SSPO_BOVIN	P98167 bos taurus
c 44	102	2.5	2768	1 THYG_RAT	P06882 rattus norv
c 45	102	2.5	5376	1 ZAN_MOUSE	O88799 mus musculu

ALIGNMENTS

RESULT 1

MTL5_MOUSE

ID Q9WTJ6; STANDARD; PRT; 295 AA.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Testin (Metallothionein-like 5, testis-specific) (testis-specific

DE metallothionein-like protein).

GN MTL5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ICR; TISSUE=Testis;

RX MEDLINE=99208669; PubMed=10191092;

RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;

RT "A novel testis-specific metallothionein-like protein, tesmin, is an

RT early marker of male germ cell differentiation.";

RL Genomics 57:130-136(1999).

CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.

CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES

CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED

CC PROGRESSIVELY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; U77383; AAD24667.1; -

DR EMBL; U67176; AAD24666.1; -

DR MGD; MGI:1340029; Mtl5.

DR InterPro; IPR005172; CXC.

Pfam; PF03638; CXC; 1.

KW Spermatogenesis.

```
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Alignment Scores:
Pred. No.: 2,9e-119 Length: 295
Score: 1599.00 Matches: 295
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.90% Indels: 0
DB: 1 Gaps: 0

US-09-743-237-1 (1-2241) x MTL5_MOUSE (1-295)
QY 651 ATGTTGATTGTCAGCTGAAGAGGCGCCGAGATGCTCTCATAGACAACCTGTGGCGCG 710
Db 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 711 AGGGAGCTCAAGCGCTCCATCTGCTCTCAGTAGATGACGAGCAGTTTCCTCAG 770
Db 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAsnSerSerPheProGln 40
QY 771 TCAGAGCTCCCTAAGCCCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACACGCGAAG 830
Db 41 SerGluLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 831 TTAATCTCATCACACAGGTTGATAATGAGCTCTCCCATCAGCTGTGCAATGGGCTGCC 890
Db 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 891 TTTCCCTCTGACCTGCTCTGCAAGGCGCCCAAAATAACTCTGCTGGGTACTGTGAC 950
Db 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGGGGACTTCTGCAACAGCTGCAGCTCAACAACCTCGCCATCAGCTC 1010
Db 101 CysPheSerSerGlyAspPheCysAsnSerCysSerCysAsnAsnLeuArgHisGluLeu 120
QY 1011 GAGCGCTTCAAGCCATGAAGCGGTCTTGATAGAAATCTGAGCTTTCCAAACCAAAA 1070
Db 121 GluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheGlnProLys 140
QY 1071 ATGGGAAGCGCTCTGGGAGCTCTAAACTTCACACACAGGCTGCAACTGTAAG 1130
Db 141 MetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGlyCysAsnCysLys 160
QY 1131 CGCTCAGGCTGCTGAAGAACTACTGTGAGTGTATGAGGCAAAATCATGTGTTCTTCC 1190
Db 161 ArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIleMetCysSerSer 180
QY 1191 ATTTCCAAATGATGCTTGGCAAAACTATGAAGAACTCCAGAACTGCAAAATGCTGATG 1250
Db 181 IleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArgLysMetLeuMet 200
QY 1251 AGCACACCCCTACATGAGCTGGGAGCTTTGAGAGCAGCATTTATTTGCCCGAGCC 1310
Db 201 SerThrProHisTyrMetGluProGlyAspPheGluSerSerHisTyrLeuSerProAla 220
QY 1311 AAGTTCTCAGGACCTCCAAAACACTGAGAAAAATAGCAGGCTTCTCCTGTATCTCTCGG 1370
Db 221 LysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSerCysIleSerTrp 240
QY 1371 GAAGTAGTGGAGCCACATGTGCTGCTGCTGGCCCGAGGCTGAGGAAGCAGCAGGAG 1430
Db 241 GluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGluAlaGluGlnGlu 260
QY 1431 CACTGTTCCCAAGCTGCTGAGCAGATGATCCTGGAGGAGTTTGGAGGCTGCTGTCG 1490
Db 261 HisCysSerProSerLeuAlaGluGlnMetIleLeuLeuGluPheGlyArgCysLeuSer 280
QY 1491 CAGATTCTCCACATCGAGTTCAAGTCCCAAGGGCTGAAATTTAG 1535
Db 281 GlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295

RESULT 2
MTL5_HUMAN
ID MTL5_HUMAN STANDARD; PRT; 299 AA.
AC Q9V4I5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).
DE MTL5.
GN MTL5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=9208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL: U86074; AAD24568.1; -.
CC Genew: HGNC:7446; MTL5.
CC MIM: 604374; -.
CC GO: GO:0005505; P:heavy metal binding activity; TAS.
CC GO: GO:0006875; P:metal ion homeostasis; TAS.
CC GO: GO:0009412; P:response to heavy metal; TAS.
CC GO: GO:0007283; P:spermatogenesis; TAS.
CC InterPro: IPR005172; CXC.
CC Pfam: PF03638; CXC; 1.
CC Spermatogenesis.
KW Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH.
SQ SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;

Alignment Scores:
Pred. No.: 1,33e-91 Length: 299
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 31.21% Indels: 4
DB: 1 Gaps: 1

US-09-743-237-1 (1-2241) x MTL5_HUMAN (1-299)
QY 651 ATGTTGATTGTCAGCTGAAGAGGCGCCGAGATGCTCTCATAGACAACCTGTGGCGCG 710
Db 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
QY 711 AGGGAGCTCAAGCGCTCCATCTGCTCTCAGTAGATGACGAGCAGTTTCCTCAG 770
Db 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGln 40
QY 771 TCAGAGCTCCCTAAGCCCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACACGCGAAG 830
Db 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
QY 831 TTAATCTCATCACACAGGTTGATAATGAGCTCTCCCATCAGCTGTGCAATGGGCTGCC 890
Db 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
QY 891 TTTCCCTCTGACCTGCTCTGCAAGGCGCCCAAAATAACTCTGCTGGGTACTGTGAC 950
Db 891 TTTCCCTCTGACCTGCTCTGCAAGGCGCCCAAAATAACTCTGCTGGGTACTGTGAC 950
```

Db 81 pheProSerGlySerThrLeuProGlyProProLysileThrLeuAlaGlyTyCysasp 100
 QY 951 TGCTTCTCCAGCGGGACCTCTGCAACAGCTGCAGC-----TGCACACACCTG 998
 Db 101 CysPheAlaSerGlyPheCysAsnAsnCysAsnAsnCysAsnAsnLeu 120
 QY 999 CGCATGAGCTCGAGCGCTTCAAGCCATAAAGCGTGTCTTGATAGAAATCTGAAGCT 1058
 Db 121 HisHisAspIleGluArgPheLysAlaLysAlaCysLeuGlyArgAsnProGluAla 140
 QY 1059 TTCCACCAAAATATGGGAAAGCGCTCGGGAGCTGCTAACTTCGACACACCAAGG 1118
 Db 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
 QY 1119 TGCACCTGTAAGCGCTCAGCTGCCTGAAGAACTACTGTGAGTGCATGAGGCCAAATC 1178
 Db 161 CysAsnCysArgArgSerGlyCysLeuLysAsnTyrcysGluCysTyrgluAlaGlnIle 180
 QY 1179 ATGTGTTCTTCCATTTGCAATGTCATTTGCAAAACTATGAAGAAAGTCCAGAACGA 1238
 Db 181 MetCysSerSerIleCysLysCysLysLysAsnTyrgluGluSerProGluArg 200
 QY 1239 AAATGCTGATGAGCACACCCACTACATGAGCGCTGGGACTTTGAGAGCAGCCATTAT 1298
 Db 201 LysThrLeuMetSerMetProAsnTyrcysGlyGlnThrGlyGlyLeuGlySerHisTy 220
 QY 1299 TTGTCCTCCAGCGCTCAGCTGCCTGCAAACTGCAAACTGCAAACTGCAAACTGCTCC 1358
 Db 221 LeuProThrLysPheSerGlyLeuProArgPheSerHisaspArgProSerSer 240
 QY 1359 TGTATCTCTCGGAAGTAGTGGAGGCGACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1418
 Db 241 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuAlaGlnGlyGluGlu 260
 QY 1419 GCAGACGAGGACACTGTTCCCAAGCTGGCTGAGCAGATGCTGAGGAGTTGGA 1478
 Db 261 AlaGluGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluPheGly 280
 QY 1479 AGTGCTGCTGCGCAGACTTCTCCACATCGAGTTCAAGTCCAAAGGGCTGAAATTTGAG 1535
 Db 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 3
 DRPL_RAT STANDARD; PRT; 1183 AA.
 AC P54258;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).
 GN DRPLA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cerebellum, and Striatum;
 RX MEDLINE=97317138; PubMed=9173996;
 RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
 RA Ashworth R.G., Ross C.A.;
 RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
 RT homologue";
 RL Neurobiol. Dis. 2:129-138(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
 RX MEDLINE=96081227; PubMed=8541849;
 RA Schmitt I., Epplen J.T., Riess O.;
 RT "Predominant neuronal expression of the gene responsible for
 RT dentatorubral-pallidolysian atrophy (DRPLA) in rat.";
 RL Hum. Mol. Genet. 4:1619-1624(1995).

CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
 CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
 CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
 CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X31777; AAA80337.1; -;
 DR EMBL; X89453; CAA61623.1; -;
 DR InterPro: IPR002951; Atrophin.
 DR Pfam: PF03154; Atrophin-1; 2.
 DR PRINTS; PR01222; ATROPHIN.
 FT DOMAIN 165 171 POLY-PRO.
 FT DOMAIN 303 306 POLY-PRO.
 FT DOMAIN 377 383 POLY-SER.
 FT DOMAIN 387 391 POLY-SER.
 FT DOMAIN 440 446 POLY-PRO.
 FT DOMAIN 477 480 POLY-HIS.
 FT DOMAIN 481 489 POLY-GLN.
 FT DOMAIN 502 505 POLY-PRO.
 FT DOMAIN 562 572 POLY-SER.
 FT DOMAIN 702 705 POLY-PRO.
 FT CONFLICT 455 455 N -> S (IN REF. 2).
 FT CONFLICT 594 594 F -> L (IN REF. 2).
 FT CONFLICT 689 689 P -> R (IN REF. 2).
 FT CONFLICT 717 717 T -> M (IN REF. 2).
 FT CONFLICT 737 737 A -> V (IN REF. 2).
 FT CONFLICT 965 965 MISSING (IN REF. 2).
 SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;
 Alignment Scores:
 Pred. No.: 0.00188 Length: 1183
 Score: 144.00 Matches: 140
 Percent Similarity: 27.98% Conservative: 48
 Best Local Similarity: 20.83% Mismatches: 177
 Query Match: 3.56% Indels: 308
 DB: 1 Gaps: 28
 US-09-743-237-1 (1-2241) x DRPL_RAT (1-1183)
 QY 1997 CCGACACCCCTCCCATATCTCTGTGCACAGGCTGCCTGGAAGAAACCCCTGATGCCA 1938
 Db 171 ProAspSerIleProArgGlnProGluSerGlyPheGlu---ProHisProSerValPro 189
 QY 1937 CCGTACTTGTATGTGATTTGGGCGAGAGTCCACCTGGAAGAGAGCGCTAGCTCTCTAACTTG 1878
 Db 190 ProThrGlyTyrr----- 193
 QY 1877 AGCGTGGCTCAATAACTTGAATGAGGTACTACTAAGCAATTGAAAGCAACACATA 1818
 Db 194 -----HisAlaProMetGluProProThrSerArgLeuPheGlnGlyProProPro 210
 QY 1817 ACAACAGTAGAATAACACCCCTCTTTTAT-----CCAGGCGAGGAGTCCCGCAGA 1767
 Db 211 GlyAlaProProHisProGlnLeuTyProGlySerAlaGlyGlyValLeuSer 230
 QY 1766 GGGCTCCCGGAGGAGGAGGAGGAGGAGTGT-----CCATCCCCAGGG 1725
 Db 231 GlyProProMetGlyProLysGlyGlyAlaAlaAlaSerSerValGlyProProSerG 250
 QY 1724 TGGCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665
 Db 250 yGlyLys-----GlnHisProProProThrThrProIleProIleSerSe 265
 QY 1664 A----- 1664
 Db 265 rSerGlyAlaSerGlyAlaProProAlaLysProProAsnThrProValGlyAlaGlyAs 285

QY 1664 ----- 1664
Db 285 nLeuProSerAlaProProAlaThrPheProHisValThrProAsnLeuProProPr 305
QY 1663 -----CTCAGACCT-----CACAGCTGACCTGCCATGATGCCGGCC 1626
Db 305 oProAlaLeuArgProLeuAsnAsnAlaSerProGlyMetGlyAlaGlnPr 325
QY 1625 TCTG-----CTCCTCCC 1614
Db 325 oileProGlyHisLeuProSerProHisAlaMetGlyGlnGlyMetSerGlyLeuProPr 345
QY 1613 TGGG-----CACCAGCTTCCTCGGTGCAGATT 1587
Db 345 oglyProGlyLysGlyProThrLeuAlaProSerProHisProLeuPro-ProAlaSerS 365
QY 1586 CCCAGGCTGAGGCTGCCACAGGATWCCCTTTACCAGCTTGACGCTACTCAATTTT 1527
Db 365 erSerAlaProGlyProProMetArgTyrProTyrSerSerCysSerSerVala 385
QY 1526 CAGCCCTGGACTTGAACCTGCATGTGAGAATCTGCGACAGGACCTTCCAAACTCT- 1468
Db 385 laAla-----SerSerSerSerAlaAlaThrSerGlnTyrProA 399
QY 1468 ----- 1468
Db 399 laSerGlnThrLeuProSerTyrProHisSerPheProProProThrSerMetSerValS 419
QY 1467 -----CCAGGATCACTGCTCAGCAAGCTTGGGAACAGTCT----- 1429
Db 419 erAsnGlnProProLysTyrThrGlnProSerLeuProSerGlnAlaValTyrSerGlnG 439
QY 1428 -----CCTGCTCTGCTCCT-----CACCCCTG 1407
Db 439 lyProProProProProProTyrGlyArgLeuLeuProAsnAsnAsnThrHisProG 459
QY 1406 GGCCA-----GCAGCAGGCACATGTGGCTCCACTACTTCCAGGAGATACA 1359
Db 459 lyProPheProProThrGlyGlyGlnSerThrAlaHisProProAlaProAlaHisHis 479
QY 1358 GGAGAAGCCCTGCTATTTTCTCAGTTTGGAGGCTCTGAGAACTTGGCTGGGGACA 1299
Db 479 lshGlnGln----- 482
QY 1298 ATAATGGTGTCTCAAGTCCCCAG----- 1273
Db 483 -----GlnGlnGlnProGlnProGlnProGlnProGlnHisHisGlyA 499
QY 1272 -----GCTCCATGTAGTGGGTGCTCATCAGCAATTTTCTGCTGCACTTCTTC 1221
Db 499 snSerGlyProProProProGlyAla----- 507
QY 1220 ATAGTTTTTGAAGCAATGCATTTGCAAAATGGAAGAACACACATGATTTGGCTCATAGCA 1161
Db 508 -----TyrProHisProL 512
QY 1160 CTCACAGTAGTCTTCA-----GCCAGCC 1137
Db 512 euGluSerSerAsnSerHisHisAlaHisProTyrAsnMetSerProSerLeuGlySerL 532
QY 1136 TCAGCGCTTACAGT-----CACCTTGTGCTGCTCAAA 1102
Db 532 euArgProTyrProProGlyProAlaHisLeuProProSerHisGlyGlnValSerTyrS 552
QY 1101 GTTTAGCAGCTCCAGAGCGCTTCCCATTTTGGTTGGAAAGCTTCAGGATTTCTAT 1042
Db 552 erGlnAlaGlyProAsnGlyProProValSerSerSerSerSerSerSerSerS 572
QY 1041 CAAGACACGCTTTATGCTTTGAAGCGCTCGAGCTCATGCGCAGGTGTGTTCAGCTGC 982
Db 572 erGlnAlaAlaTyrSerCysSerHisProSerSer----- 584

QY 981 AGCTTCTGCAGAAAGTCCCGCTGGAGAAGCAGTCACTACAGTACCCAGACAGAGTTATTTGG 922
Db 584 ----- 584
QY 921 GTGGCCCTTGACAGCAGGTCCAGAGGAGAAAGGACCCCATTTGACAGCTGATGGAGAG 862
Db 585 -----GlnGlyProGlnGlyAlaSerTyrProPhe-----P 595
QY 861 CTCATTATCAACCTGTGTGATGAGATTTAACTTCCTGGTACTGGCAGAACTCTTCCCA 802
Db 595 roProVal-ProProile-----ThrThrSerSerAlaThrLeuSerThrValle--- 611
QY 801 TAAAGTTGTTCATGGCTTAGGAGCTCTGAGGAACTGCTGAGGAACTGCTGCTGCTACT 742
Db 612 -----AlaThrValAlaSerSerProAlaGlyTyrLysThr 623
QY 741 GAGGAAGCAGATGGAGCGCTTTGAGCTCCCTCGCGCCACAGTTGCTATGACAGACATCT 682
Db 624 AlaSerProProGly-----ProProGlnTyrSerLysArgAlaProSerPro 639
QY 681 GGGCGCTCCCTTTCAGCTGACAAATCAGCATGGGCTGGAGTCTTTCTTCCGAGGCGCAG 622
Db 640 GlySerTyrLysThrAla-----ThrProProGlyTyrLysProGlySerProSer 657
QY 621 TGGAGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 562
Db 658 PheArgThrGlyThrProGlyTyrArgGlyThr-----SerProProAla 673
QY 561 GGGCTGCGAAAGCTGCTGCTTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 511
Db 674 GlyProGlyThrPheLys-----ProGlySerProThrValGlyProGlyProLeuProProA 693
QY 510 CT-----TCTTGTATTTCAACACTT 491
Db 693 laGlyProSerSerLeuSerSerLeu 701
RESULT 4
HIC1_MOUSE
ID HIC1_MOUSE STANDARD; PRT; 892 AA.
AC Q9RLY5; Q9RLY6; Q9R2B0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypermethylated in cancer 1 protein (Hic-1).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC STRAIN=129/Sv, and Swiss Webster; TISSUE=Embryo;
RX MEDLINE=99172081; PubMed=10072440;
RA Grimm C., Spoerle R., Schmid T.E., Adler I.-D., Adamski J.,
RA Schughart K., Graw J.;
RT "Isolation and embryonic expression of the novel mouse gene Hic1, the
RT homologue of Hic1, a candidate gene for the Miller-Dieker syndrome.";
RL Hum. Mol. Genet. 8:697-710(1999).
RN [2]
RP SEQUENCE FROM N.A., DISEASE, AND TISSUE SPECIFICITY.
RX MEDLINE=20122251; PubMed=10655551;
RA Carter M.G., Johns M.A., Zeng X., Zhou L., Zink M.C., Mankowski J.L.,
RA Donovan D.M., Baylin S.B.;
RT "Mice deficient in the candidate tumor suppressor gene Hic1 exhibit
RT developmental defects of structures affected in the Miller-Dieker
RT syndrome.";
RL Hum. Mol. Genet. 9:413-419(2000).
RN [3]
RP SEQUENCE OF 179-338 FROM N.A.
RC STRAIN=129/Sv; TISSUE=Liver;
RX MEDLINE=99297610; PubMed=10371200;
RA Guerardel C., Deltour S., Leprince D.;
RT "Evolutionary divergence in the broad complex, trantrack and bric a

brac/poxviruses and zinc finger domain from the candidate tumor suppressor gene hypermethylated in cancer.";

FEBS Lett. 451:253-256(1999).

-I- FUNCTION: Putative transcription factor. May act as a tumor suppressor. May be involved in development of head, face, limbs and ventral body wall.

-I- SUBUNIT: Interacts with CtBP (By similarity).

-I- SUBCELLULAR LOCATION: Nuclear (Probable).

-I- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=1;

Comment=A number of isoforms may be produced;

Name=1;

CC IsoId=Q9RLY5-1; Sequence=Displayed;

CC TISSUE SPECIFICITY: Ubiquitously expressed with highest levels in heart and lung.

CC -I- DEVELOPMENTAL STAGE: Expression is first detected in the embryo after 9 dpc. In the embryo, expression is found in restricted regions of somite derivatives, limb anlagen and cranio-facial mesenchyme. In the fetus, it is additionally expressed in mesenchyme apposed to precartilaginous condensations, at many interfaces to budding epithelia of inner organs, and weakly in muscles.

CC -I- DISEASE: Defects in HIC1 are the cause of perinatal death with serious developmental anomalies, including acrania, exencephaly, cleft palate, omphalocele, craniofacial and limb anomalies.

CC -I- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.

CC -I- SIMILARITY: Contains 1 BTB/POZ domain.

CC -I- SIMILARITY: Contains 5 C2H2-type zinc fingers.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF036334; AAD30654.1; -

DR EMBL; AF036582; AAD30655.1; -

DR EMBL; AJ132691; CAB44493.1; -

DR HSP; P08046; IAH.

DR MGI; MGI:1338010; Hic1.

DR InterPro; IPR000210; BTB_POZ.

DR InterPro; IPR007087; Znf_C2H2.

DR Pfam; PF00651; BTB; 1

DR Pfam; PF000936; zf-C2H2; 5.

DR ProDom; PD0000003; Znf_C2H2; 1.

DR SMART; SM00225; BTB; 1.

DR SMART; SM00355; Znf_C2H2; 5.

DR PROSITE; PS50097; BTB; 1.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.

DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.

KW Multigene family; Nuclear protein; Transcription regulation;

KW DNA-binding; Zinc-finger; Metal-binding; Repeat;

KW Developmental protein; Alternative splicing.

FT DOMAIN 1 172 ARG/PRO/GLY-RICH.

FT DOMAIN 206 269

FT DOMAIN 271 278

FT DOMAIN 319 326 POLY-ALA.

FT DOMAIN 354 358 POLY-GLY.

FT ZN_FING 596 623 POLY-PRO.

FT ZN_FING 666 693 C2H2-TYPE 1.

FT ZN_FING 694 721 C2H2-TYPE 2.

FT ZN_FING 722 749 C2H2-TYPE 3.

FT ZN_FING 750 777 C2H2-TYPE 4.

FT ZN_FING 772 777 C2H2-TYPE 5.

FT SITE 400 404 BINDING TO CTBP.

FT SITE 242 242 N -> S (IN REF. 1; AAD30654).

FT CONFLICT 242 242

CC SEQUENCE 892 AA; 94320 MW; BBD16AA051995740 CRC64;

Alignment Scores: 0.0516 Length: 892

Pred. No.: 125/50 Matches: 173

Score:


```
Db 558 toAsnGlyPro---ProVal-----563
QY 1029 TTATGGCTTTGAAGCGCTCGAGCTCATCGCGCAGGCTTGTTCAGCTGCAGCTGTTCGAGA 970
Db 564 -----SerSerSer-----AsnS 569
QY 969 AGTCCCGCTGGAGAAGCAGTACAGTACCAGACAGAGTTATTTTGGTGGCGCCCTTCG- 911
Db 569 erSerSerThrSerGlnGlySerTyr-----ProCysS 581
QY 910 -----AGACAGGCTCCAGAGGGAAGAGCCAGCCCACTGACAGCTGATGGA 865
Db 581 erHisProSerProSerGlnGlyProGlnGlyAlaProTyrProPhe-----596
QY 864 GAGCTCCATTATCAACCTGTGTGATGAGATTAACTCGCTGGTACTGGCAGAGTCTTC 805
Db 597 --ProProValProThr-Val-----ThrThrSerSerAlaThrLeuSerThrValIle 613
QY 804 CCACTAAAGTGTCTATTCGTTAGGAGCTCTGACTGAGGGAAGTCTCTGGTTCATCGT 745
Db 614 -----AlaThrValAlaSerSerProAlaGlyTyrLys 624
QY 744 ACTGAGGAAGCAGATGAGCGCTTTGAGCTCCCTCGCGCCACAGTTGCTATGCAGACA 685
Db 625 ThrAlaSerProProGly-----ProProTyrGlyLysArgAlaProSer 640
QY 684 TCTGGCGCTCTCTTTCAGCTGACAAATCACCATGGGCTGGAGTCTTTCTTCGAGGCG 625
Db 641 ProGlyAlaTyrLysThrAla-----ThrProProGlyTyrLysProGlySerProPro 658
QY 624 AGCTGGAGGCTCTCTCTGCTGCTGGGATGATGGGAATTCGAACAGATTCTCTGAGACA 565
Db 659 SerPheArgThrGlyThrProProGlyTyrArgGlyThr-----SerProPro 674
QY 564 GAGGGGCTGGAAGCTCGCTCTCAGGCTGCGC-----CCTGGCACACTACCAC 514
Db 675 AlaGlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuProP 694
QY 513 CTGCT 509
Db 694 toAla 695
```

RESULT 6

```
HA44_CHICK
ID HA44_CHICK STANDARD; PRT; 309 AA.
AC P1727;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A4 (Chox-1.4).
GN HOXA4 OR CHOX-1.4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245562; PubMed=1970866;
RA Sasaki H., Yokoyama E., Kuroiwa A.;
RT "Specific DNA binding of the two chicken Deformed family homeodomain
RL proteins, Chox-1.4 and Chox-a.,"
RL Nucleic Acids Res. 18:1739-1747(1990).
RN [2]
RP SEQUENCE OF 207-273 FROM N.A.
RC STRAIN=Conet Hubbard hybrid;
RA Medline=90326535; PubMed=1973835;
RT "Isolation and analysis of chick homeobox cDNA clones.,"
RL Nucleic Acids Res. 18:3999-3999(1990).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
```

```
CC -!- FUNCTION: BINDS TO SITES IN THE 5'-FLANKING SEQUENCE OF ITS CODING
CC REGION WITH VARIOUS AFFINITIES. THE CONSENSUS SEQUENCES OF THE
CC HIGH AND LOW AFFINITY BINDING SITES ARE TAATGA(C/G) AND CTAATTTT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: THE PROLINE STRETCH WORKS AS A PART OF THE
CC TRANSCRIPTIONAL ACTIVATION DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.
CC "DEFORMED" SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52670; CAA36896.1; -
CC EMBL; X52747; CAB57949.1; -
CC PIR; S09257; S09257.
CC HSP; P02833; 9ANT.
CC TRANSFAC; T00128; -
CC InterPro; IPR001827; Antennapedia.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00025; ANTENNAPEDIA.
CC PRINTS; PR00024; HOMEBOX.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC PROSITE; PS00032; ANTENNAPEDIA; 1.
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;
CC Transcription regulation.
CC FT DOMAIN 38 148 PRO-RICH
CC FT SITE 188 193 ANTP-TYPE HEXAPEPTIDE.
CC FT DNA_BIND 209 268 HOMEBOX.
CC SQ SEQUENCE 309 AA; 33478 MW; 8898A41502319341 CRC64;
Alignment Scores:
Pred. No.: 0.169 Length: 309
Score: 117.50 Matches: 58
Percent Similarity: 36.36% Conservative: 14
Best Local Similarity: 29.29% Mismatches: 81
Query Match: 2.90% Indels: 45
DB: 1 Gaps: 9
US-09-743-237-1 (1-2241) x HA44_CHICK (1-309)
QY 1829 AGCAACACATAACAACAAAGTAGAAAATCACCCCTCTTTTATGAGGCGGAGTCCCC 1770
Db 24 ThrGlnHisSerGlySerAlaGlySerAlaSerTyrHisProHisProHisPro 43
QY 1769 AGAGGGCTCCCGGAGGACAGGCGGCGAGTGTCCATCCCGAGGCTCTCGCTGCTT 1710
Db 44 HisAlaProProPro-----ProProProProProProHisLe 57
QY 1709 GCACCTTTAGGATACCTTGAGTAGCTGCGCAGTACATGCAGATCACATCAGCTCACAG 1650
Db 57 uHisAlaAlaHisPro-----GlyProAlaLeuProGluTyrPheProArgArg 75
QY 1649 CTGACCTGGCCATGATGCGGCGCTCTCTCTCCCTGGGACACAGCTTCCTCGGTGAG 1590
Db 75 gGluProGlyTyrGlnAla--ProAlaAlaProProGlyProProGlyProPro----- 92
QY 1589 ATTCGAGGCGCTGAGGCTTGGCCACAGGATTCCTCCCTTTACCAGCTTGCAGCGTACTCAAT 1530
Db 93 -----ProGluAlaLeuTyrProAlaGlnAlaProSerTyrProGlnAlaProTyrSer-T 111
QY 1529 TTTCAGCCCTTGGACTTGAACCTCGATGCTGGAGAATCTCGGAGACAGGACCTTCCAACTC 1470
Db 111 yrSerSer-----AlaGlySerAlaAlaProGlyProGluGlnProP 125
```


QY 1062 -----GGAGCTTCAGGATTTCTATCAAGACAGCGCTTATGCG 1023
 Db 693 rpAspValHisGlyAsnGlyThrGlnGlnThrPheTyrGlnAspProSerValLeuT 713
 QY 1022 TTGTAAGCGCTCGAGCTCATGCGGAGGT 994
 Db 713 yrIleSerLeuHisArgHisAspGly 722

RESULT 8
 RECK_HUMAN STANDARD; PRT; 971 AA.
 AC O95980; Q8WX37;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Reversion-inducing cysteine-rich protein with kazal motifs precursor
 DE (hRECK) (Suppressor of tumorigenicity 15) (ST15).
 GN RECK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND N-GLYCOSYLATION.
 RC TISSUE=Fibroblast;
 RX MEDLINE=99007295; PubMed=9789069;
 RA Takahashi C., Sheng Z., Horan T.P., Kitayama H., Maki M., Hitomi K.,
 RA Kitaura Y., Takai S., Sasahara R.M., Horimoto A., Ikawa Y.,
 RA Ratkin B.J., Arai K., Noda M.;
 RT Regulation of matrix metalloproteinase-9 and inhibition of tumor
 RT invasion by the membrane-anchored glycoprotein RECK.;
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13221-13226(1998).
 RN [2]
 RP SEQUENCE OF 363-971 FROM N.A.
 RA Kimberley A.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Negatively regulates matrix metalloproteinase-9 (MMP-9)
 CC by suppressing MMP-9 secretion and by direct inhibition of its
 CC enzymatic activity. RECK down-regulation by oncogenic signals may
 CC facilitate tumor invasion and metastasis. Appears to also
 CC regulate MMP-2 and MT1-MMP, which are involved in cancer
 CC progression.
 CC -1- SUBUNIT: Interacts with MMP-9.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: Expressed in various tissues and untransformed
 CC cells. It is undetectable in tumor-derived cell lines and
 CC oncogenically transformed cells.
 CC -1- SIMILARITY: Contains 3 Kazal-like domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D50406; BAA34060.1; -;
 DR EMBL; AL158830; CAD13384.1; -;
 DR Genbank; HGNC:11345; RECK.
 DR MIM; 605227; -;
 DR HSSP; P80424; IAN1.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0008191; F:metalloendopeptidase inhibitor activity; TAS.
 DR InterPro; IPR0023350; kazal.
 DR Pfam; PF00050; kazal; 2.
 DR SMART; SM00280; KAZAL; 3.
 DR PROSITE; PS00282; KAZAL; 1.
 KW Signal; Glycoprotein; GPI-anchor; Serine protease inhibitor;
 KW Membrane; Anti-oncogene; Repeat.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 942 REVERSION-INDUCING CYSTEINE-RICH PROTEIN

FT PROPEP 943 971
 FT LIPID 942 971
 FT DOMAIN 632 677
 FT DOMAIN 708 750
 FT DOMAIN 753 787
 FT DOMAIN 37 338
 FT REPEAT 37 84
 FT REPEAT 104 141
 FT REPEAT 151 197
 FT REPEAT 216 263
 FT REPEAT 292 338
 FT REPEAT 633 654
 FT DISULFID 643 677
 FT CARBOHYD 39 39
 FT CARBOHYD 86 86
 FT CARBOHYD 200 200
 FT CARBOHYD 297 297
 FT CARBOHYD 352 352
 FT SEQUENCE 971 AA; 106456 MW; 173D47D6AEE6F834 CRC64;
 Alignment Scores:
 Pred. No.: 0.228 Length: 971
 Score: 117.50 Matches: 84
 Percent Similarity: 30.84% Conservative: 52
 Best Local Similarity: 19.05% Mismatches: 141
 Query Match: 2.93% Indels: 164
 DB: 1 Gaps: 23

US-09-743-237-1 (1-2241) x RECK_HUMAN (1-971)
 QY 660 TGTGAGTGAAGGAGGCGCCGAGATGCTCTGCATAGACAACCTGTGCGCGAGGAGCTC 719
 Db 422 CysHisSerLysSerArgGlySerIleIleLysLysSerAspCys----- 436
 QY 720 AAAGCGCTCCATCTGCTCTCCTCAGTACGATGACGAGAGCAGGTTCCCTCAGTCAGAGCTC 779
 Db 437 -----ValGluIleLeuLysLysCysGlyAspGlnAsnLysPheProGluAspHisThr 454
 QY 780 CCTAAGCCCAATCACAACATTTAGTGGGA-----AGACTTCTGCCAGTA 821
 Db 455 AlaGluSerIleCysGluLeuLeuSerProThrAspAspLeuLysAsnCysIleProLeu 474
 QY 822 -----CCAGCGAAGTGA-----AATCTCATCACACAGGTTGATATGAGCT 863
 Db 475 AspThrTyrLeuArgProSerThrLeuGlyAsnIleValGluValThrHisProCys 494
 QY 864 CTCCCATCA-----GCTGTCATGCGGCTGCCTTTCCCTCT 899
 Db 495 AsnProAsnProCysProAlaAsnGluLeuCysGluValAsnArgLysGlyCysProSer 514
 QY 900 GGA---CCTGCTCTGCAAGGCGCCACCAAAATACTCTGTCTGGGTACTGTGACTGCTTC 956
 Db 515 GlyAspProCysLeu-----ProTyrPheCysValGlnGly---CysLysLeuGly 530
 QY 957 TCCAGCGGGGACTTCTGCAACAGCTGCAGCTGCAACAACCTGCGCCATGAGCTCGAGCGC 1016
 Db 531 GluAlaSerAspPhe----- 535
 QY 1017 TTCAAAGCCATAAAGCGGTGCTTGTATAGAAAATCTTCAAGCTTTCCAAACCAAAATGGGG 1076
 Db 536 -----IleValArgGlnGlyThrLeuIleGlnValProSerSer 548
 QY 1077 AAAGCGCTCTGGAGCTGCTTAACCTTCGACACAGCAAGGCTGCAACTGTAGCCCTCA 1136
 Db 549 AlaGlyGluValGlyCysTyrLysIle-----CysSerCysGlyGlnSer 563
 QY 1137 GCGTGCCTGAAG----- 1148
 Db 564 GlyLeuLeuGluAsnCysMetGluMetHisCysIleAspLeuGlnLysSerCysIleVal 583
 QY 1149 -----AACTACTGTGAGTGC 1163

Db 584 GlyGlyLysArgLysSerHisGlyThrSerPheSerIleAspCysAsnValCysSerCys 603
 QY 1164 TATGAGGCCAAATCATGTGTTCTTCC----- 1190
 Db 604 PheAlaGlyAsnLeuValCysSerThrArgLysCysLeuSerGluHisSerSerGluAsp 623
 QY 1191 -----ATTGCGAAATGCATT-----GCTTGC 1211
 Db 624 AspArgArgThrPheThrGlyLeuProCysAsnGlnPheValProValCys 643
 QY 1212 -----AAAACATGAGAAAGTCCAGAACGAAATGCTGATGAGCACACCC 1259
 Db 644 GlyGlnAsnGlyArgThrTyrosProSerAlaCysIleAlaArgCysValGlyLeuGlnAsp 663
 QY 1260 CACTACATGGAGCGCTGGGACTTTGAGAGCAGCCATTATTTGCTCCAGCCCAAGTTCTCA 1319
 Db 664 HisGlnPheGluPheGlySerCysMetSerLys-----AspProCys----- 677
 QY 1320 GGACCTCCAAACTGAGAAAATAGGAGCGCC----- 1352
 Db 678 AsnProAsnProCysGlnLysAsnGlnArgCysIleProLysProGlnValCysLeuThr 697
 QY 1353 -----TTCCTCTGTATCTCTGGGAAGTAGTGGAGCCACATGTCCTGCCTG 1400
 Db 698 ThrPheAspLysPheGlyCysSerGlnTyrgLysValProArgGlnLeuAlaCysAsp 717
 QY 1401 CTGGCCCGAGGTGAGGAAGCAGCAGCAGTGTCTCCCA-----AGCTTG 1448
 Db 718 GlnValGlnAspProValCysAspThrAspHisMetGluHisAsnAsnLeuCysThrLeu 737
 QY 1449 GCTGAGCAGATGATCTGAGGAGGTTGGAGGTGCTGTCGAGATTCCTCCACATCGAG 1508
 Db 738 TyrGln-ArgGlyLysSerLeuSerTyrgLysProCysGlnProPheCysArgAlaTh 757
 QY 1509 TTCA-----ACTCAAGGGGTGAAATGAGTAGCTGCAAGCTGT----- 1551
 Db 757 rGluProValCysGlyHisAsnGlyGluThrTyrgSerValCysAlaAlaTyrgSerAs 777
 QY 1552 -----AAAGGGGAATGCTGTGGCAAGCCTCAGCCCTGGGAATCTG 1592
 Db 777 pArgValAlaValAspTyrgLysAspCys-----GlnAlaValGlyValLe 793
 QY 1593 CACCGAGGAGCTGTGCTCCAGGAGGAGCAGAGCCCGCATCATGCGCAGGTGAGCTG 1652
 Db 793 uSerGluHisSerSerVal-----AlaGluCysAla-----SerValLysCy 807
 QY 1653 T 1653
 Db 807 s 807

RESULT 9

ID BAR3_CHITE STANDARD; PRT; 1700 AA.
 AC Q03376;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Balbiani ring protein 3 precursor.
 GN BR3
 OS Chironomus tentans (Widge).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomidae;
 OC Chironomidae; Chironominae; Chironomus.
 OX NCBI_TaxID=7153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RX MEDLINE=90172404; PubMed=1689777;
 RA Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
 RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
 RT repetitive structure split by many introns.";
 RL J. Mol. Biol. 211:331-349(1990).

CC -!- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
 CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
 CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
 CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
 CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
 CC SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: SALIVARY GLAND.
 CC -!- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X52263; CAA36506.1; -;
 CC PIR; S08167; S08167.
 CC HSSP; P15358; 1SKZ.
 CC InterPro; IPR004153; CXCXC_repeat.
 CC Pfam; PF03128; CXCXC; 71.
 CC Repeat; Signal.
 CC SIGNAL 1 20 POTENTIAL.
 CC FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.
 CC FT SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;
 SQ
 Alignment Scores:
 Pred No.: 0.265 Length: 1700
 Score: 117.50 Matches: 100
 Percent Similarity: 31.11% Conservative: 49
 Best Local Similarity: 20.88% Mismatches: 147
 Query Match: 2.93% Indels: 183
 DB: 1 Gaps: 29
 US-09-743-237-1 (1-2241) x BAR3_CHITE (1-1700)
 QY 579 TGTGTCAAGTTCCTCA---TCATCCAGGAGCAGAGGAGGCTCCAGC----- 623
 Db 152 CysCysGlyCysProValAsnMetGlnGluProAlaAspGlyCysThrLysProLeuIle 171
 QY 624 -----TGCCTCGGAAGAAGAC----- 641
 Db 172 TrpAspLysValAspCysArgCysGluCysProLeuLysLysAspCysGlyLysAsnArg 191
 QY 642 -----TCCAGCCCATGTTGTTGTCAGCTGAAAGGA-----GGCGCC 680
 Db 192 AspTrpSerAspSerSerCysGluCysLysGlyAspGlyLysCysGlnGlySer 211
 QY 681 CAGATGCTCTCATGACAACTGTGGCGCGAGGAGCTCAAAGCGCTCCATCTGCTTCT 740
 Db 212 LysIleTrpCysLysAsnAsnCys--ArgCysIleCysProThrAlaGluProAlaGly 231
 QY 741 CAGTACGATGACAGCAGCTTTCCTCAGTCAGAGCTCCCTAAGCCCAATGCAACTTTA 800
 Db 231 LysSerAlaProLeuLysTrpAsp-----AspAspLysCys 244
 QY 801 GTGGGAAGACTTCTCCAGTACACGAGTAAATCTCAT-----CACACAGGTGAT 854
 Db 244 erCysAlaCysProAlaLysMetGluLysLysGluLysCysValGluSerGlyLysI 264
 QY 855 AATGAGCTCTCCATCAGCTGTCATGAGGGGCTGCTTCCCTCTGGACCTGCTTCTGCAA 914
 Db 264 leTrp--AsnProAsnThrCysGluCysGlyCysAlaGlnLeu--AsnCysProAsp 282
 QY 915 GGGCCACCAAAATAACTCTGCTGGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
 Db 282 snLysLysAlaAsnLysGlu--ThrCysGlnCysGluCysLysGluValLysLysCysAsn 301
 QY 960 AGCGGGGAC---TCTTCG---AACAGCTGACGCTGCAACAACTGCCCATGAGCTCGAG 1013
 Db 302 GlyGlyGlnValPheCysLysAspSerCysSerCys----- 313

QY 1014 GCCTTCAAGCCATTAAGCGCGTCTT-----GATAGAAATCCGAAGCTTTCACACA 1067
 Db 314 -----ValCysProGlyGlyAspLysThrCysThrAlaPro 327
 QY 1068 AAAATGGGAAAGCGCGTCTGGAGCTGCTAAACTTCGACACAGCAAGGGTGCACACTGT 1127
 Db 328 GlnValTyrAspGly-----ValAlaCysSerCys 337
 QY 1128 AAG-----CGCTCAGGCTGCCTG----- 1145
 Db 338 SerCysProValAsnMetClnLysProAlaAspGlyCysProArgProGlnLysTrpasp 357
 QY 1146 AAGAACTACTGT-----GAGTGTATGAGCCCAAAATCATG 1181
 Db 358 LysGluGluCysArgCysGluCysProValLysHisAspCysLysAsnGlyLysValTrp 377
 QY 1182 TGTCTTCATTTGCAATGCAATGCT----- 1208
 Db 378 AspGluThrileCysGlnCysileCysProArgAspAlaProValCysThrAlaGlyLys 397
 QY 1209 -----TCCAAAAAATATGAAGAAAGTCCAGAACGAAAGAAATG 1244
 Db 398 GluArgCysGlyGluSerCysGluCysLysCysLysCysIleAsnArgGluProLys-GluGlyCys 417
 QY 1245 CTGATGAGACACCCACTACATGAGCGCTGGGACTTTGAGAGCAGCATTTATTTGTC 1304
 Db 417 s-----AlaLysProLeuValTrpAsnGluAsnThrCysLysCys-----ValCysPr 433
 QY 1305 CCAGCCAAAGTCTCAGGACCTCAAAACTGAGAAATAATAGCAGCGCTTCTCTGTATC 1364
 Db 433 o-----AlaAspLysGln-----MetSe 439
 QY 1365 TCCTGGG----- 1371
 Db 439 rProGlyGlyCysGlySerGlyLysSerPheAsnLysLeuThrCysGlnCysGluCysAs 459
 QY 1372 -----AAGTAGTGGAGGCCACATGTGCCCTGCTGTGGCC 1406
 Db 459 pGlnSerAlaSerLysCysGlyLeuLysArgTrpAsnAlaAspThrCysLysCys----- 477
 QY 1407 CAGGTGAGGAGCAGCAGGACGACTTTCCTCCCAAGCTTGGCTGAGCAGATGCTCTG 1466
 Db 478 -----GluCysGlnProGlyMetProGluGlyCysGlyLysGlnThrTr 493
 QY 1467 GAGGAGTTGGAGGCTGCGAGATTCTCCACATCGAGTCCAAAGTCCAAAGGGGCTG 1526
 Db 493 pIleSerAspLysCysLysCysGlu-CysSerPro---ThrileThrCysGlnAlaProG 512
 QY 1527 AA-----AATTGAGTAGCGTGCAAG-----CTGGTAAAGGGGA 1559
 Db 512 lnIleLeuAspLeuAsnThrCysGluCysLysCysProValAsnMetLeuAlaGlnLysG 532
 QY 1560 ATGCTGTGGCAAGCTCAGCCCTGGGATCTGCACCGAGGAGTGGTGC----- 1610
 Db 532 luLysCysLysSerProArgGlnTrpThrAspSer---LysCysLeu-CysGluCysSer 550
 QY 1611 -----CCAGGAGGAGCAGAGCGCGCATCATGGCCAGGTCAG 1649
 Db 551 ThrThrProAlaThrCysGluGlyLysGlnThrTrpCysGlyGlu 565

RESULT 10

GSRL_HUMAN
 ID GSRL_HUMAN STANDARD; PRT; 1509 AA.
 AC Q9NZM4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Glioma tumor suppressor candidate region gene 1 protein.
 GN GLTSCR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

Alignment Scores:
 Pred. No.: 0.37 Length: 1509
 Score: 115.50 Matches: 140
 Percent Similarity: 32.03% Conservative: 56
 Best Local Similarity: 22.88% Mismatches: 217
 Query Match: 2.85% Indels: 199
 DB: 1 Gaps: 29

US-09-743-237-1 (1-2241) x GSRL_HUMAN (1-1509)

QY 2065 GTTTCTCTACTTCACGACGAGAAATACAGGAAGAAAAAGCAAAAATCCAAACAAACCC 2006
 Db 783 lIlePheValIleGlnAsnGlnLeuGlyValProProAlaSerAsnPro---AlaPro 801
 QY 2005 AGCTCTAACCCGCCCTCCCTCCATATCTCTGTACAGCGCTGGCTGGAACCCCT 1946
 Db 802 ThrAlaProGlyProProGlnProLeuArgProGlnSerGlnPro-----Pro 818
 QY 1945 GATGCCACCTGACTTGTATGTGTTGGCGAGAGTCCACCTGAAGAAGAGCTAGCTCT 1886
 Db 819 GluGlyProLeu-----ProProAlaProHisLeuProProSerSer 832
 QY 1885 CTAACCTGAGCGCTGCATATACTTAAGTGGAGGTACTACTAAGCAATTTGAAGCA 1826
 Db 833 ThrSer-----SerAlaValAlaSerSerGluThr 843
 QY 1825 AACACATACAAAGTAGAAAATCACCTCTTTTATCAGGCGAGGAGTCCCCAGAG 1766
 Db 844 SerSer-----ArgLeuProAla 849
 QY 1765 GGCCTCCCGGAGGAGGAGGCGCAGTGTCCATCCCGAGGTGGCTCTGCTGCTTGCAC 1706
 Db 850 ProThrProSerAspPheGlnLeuGlnPheProProSerGlnGly----- 864
 QY 1705 TTTAGATACCTTGTAGTAGGCTGGCCAGTACCATGACATGACATCTAGACCT----- 1655
 Db 865 -----ProHisLysSerProThrProProThr 874
 QY 1654 ---CACAGCTGACCTGGCCATGATGGCGGCTCTGCTCTCCTCCT---GGGCACCAAGCTT 1601
 Db 875 LeuHisLeuValProGluProAlaAlaProProProProProProProProProPro 894

SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 MEDLINE=20175430; PubMed=10708517;
 Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
 Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
 Scheithauer B.W., Louis D.N., Jenkins R.B.;
 "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
 region.";
 Genomics 64:44-50(2000).
 CC TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
 placenta, skeletal muscle, and pancreas, and at lower levels in
 lung, liver, and kidney.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@isb-sib.ch).

EMBL: AF182077; AAF62874.1;
 Genew: HGNC:4332; GLTSCR1.
 MIM: 605690;
 DR DOMAIN 37 45 POLY-GLY.
 FT DOMAIN 884 889 POLY-PRO.
 FT DOMAIN 1214 1225 POLY-SER.
 FT DOMAIN 1282 1286 POLY-PRO.
 FT DOMAIN 1294 1304 POLY-PRO.
 SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;

```

QY 1600 CCTGGTCCAGATCCCGAGGCTG-----AGG 1574
Db ValThrThrProPheProAlaLeuLeuProGlnProLysAlaLeuLeuGluArgPheHisGln 914
QY 1573 CTGCGCAGGCAATCCCTTACCAGCTTCACGCTACTCAATTTTCAGCCCTTGAC 1514
Db ValProSerGlyIleLeuLeuGlnAsnLysAla-GlyGlyAlaProAlaAlaProGlnTh 934
QY 1513 TTGAACCTCGATGTGGAGAATCTGGCAGAGGACCTTCCAAACTCCTCCAGGATCATGCG 1454
Db rSerThrSerLeuGlyPro-----LeuThrSerProAlaAlaSerVa 948
QY 1453 T-----CAGCCAAAGTGGGGAACAGCTGCTCCTGCTGCTTCCTCACCC 1409
Db lLeuValSerGlyGlnAlaProSerGlyThrProThrAlaProSerHisAlaProAlaPr 968
QY 1408 TGGGCCAGCAGCAGGACATGTGGCTCCACTACTCCAGGAGATACAGGAGAGGCC 1349
Db oAlaProMetAlaAlaThrGlyLeuProProLeuProAlaGluAsnLysAla----- 986
QY 1348 TGCCTATTCTTCTCAGTTTGGAGGCTCTGAGAACTTGGCTGGGACAAATAATGGCTG 1289
Db -----PheAlaSerAsnLeuProThrLeuAsnValAlaLysAlaAlaSerSerGly-- 1003
QY 1288 CTCCTAAAGTCCCGCAGGCTCCATAGTGGGTGTGCTCATCAGCAATTTTCGTTCTGGA 1229
Db -----ProGlyLysPro--SerGlyLeuGlnTyroGluSerLysLeu--SerGly 1018
QY 1228 CTT-----TCTTCATAGTTTTCGAAAGCAATGCAT 1199
Db LeuLysLysProProThrLeuGlnProSerLysGluAlaCysPheLeuGlu-----His 1036
QY 1198 TTGCAAAATGGAAGAACACATGATTGGCCCTCATAGCACTCACAGTAGTTC-----TTC 1145
Db LeuHisLysHisGlnGlySerValLeuHisProAspTyrLysThrAlaPheProSerPhe 1056
QY 1144 AGGCAGCTGAGCGCTACAGTTGCACCTTGTGTCGTCGCAAGTTTAGCAGCTCCAGCA 1085
Db -----AlaLeuHisArgLeuLeuProTyrHisValTyrGlnGlyAla 1072
QY 1084 CGGCTTTCCCACTTTTGGTTGGAAAGCTTCAGGATTTCTATCAAGACAGCGCTTTATG 1025
Db LeuProSerProSerAspTyrHisLysValAspGluGluPheGluThrValSerThrGln 1092
QY 1024 GCTTTGAAGCTCGAGCTCAGTCAGCGGAGGTGTGTCAGCTGTCGAGCTGTCGAGAGTCC 965
Db LeuLeuLysArgThrGlnAlaMetLeuAsnLysTyrArgLeuLeuLeuGluGluSer 1112
QY 965 ----- 965
Db 1113 ArgArgValSerProSerAlaGluMetValMetIleAspArgMetPheIleGlnGlu 1132
QY 964 -----CGCGTGGGAAGCAG-----TCACAGTACCCAGCAGAGTATTATT--- 926
Db LysThrThrLeuAlaLeuAspLysGlnLeuAlaLysGluLysProAspGluTyrValSer 1152
QY 925 -----TTGGGTGGCCCTTGACAGACAGCTCCAGGAGGAAAGGAGGCCCAATG 878
Db SerSerArgSerLeuGlyLeuProIleAlaAlaSerSerGluGlyHisArgLeuPro--- 1171
QY 877 ACAGCTGATGGAGAGCTCCATTATCAACCTGCTGTGATGAGATTTAACTTCGCTGGTACT 818
Db -----GlyHisGlyProLeuSerSer-----SerAlaPro 1181
QY 817 GGCAGAGCT-----CTTCCCACTAAAGTTGTCAAT-----GGC 785
Db GlyAlaSerThrGlnProProHisLeuProThrLysLeuValIleArgHisGlyGly 1201
QY 784 TTAGGGAGCTCTGACTGAGGGAACCTGCTGTGCTATCGTACTGAGGAGCAGATGGAGC 725
Db AlaGlyGlySerPro-----SerValThrTrpAla-----Arg 1212
QY 724 CTTTGGAGCTCCCTCGGCCACACATGTTGCTATGACAGACATCTCGGGCCCTCTTTTCAGC 665

```

```

Db 1213 AlaSerSerLeuSerSerSerSerSerSerSerSerSerSer-AlaAlaSerSerLeuAspAl 1232
QY 664 TGAC-----AA 659
Db 1232 aAspGluAspGlyProMetProSerArgAsnArgProProIleLysThrTyrGluAlaAr 1252
QY 658 ATCACCATTGGGCTGGAGTCTT----- 637
Db 1252 gSerArgIleGlyLeuLysLeuLysIleLysGlnGluAlaGlyLeuSerLysValValHi 1272
QY 636 -----TCTTCCGAGGCGAGCTGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 584
Db 1272 sAsnThrAlaLeuAspProValHisGlnProProProProProAlaThrLeuLysValAl 1292
QY 583 CAACAGGATTCCTGAGCCAGAGGGGCGCT 556
Db 1292 aGluProProProArgProProProPro 1301

```

RESULT 11

CD93_RAT

ID CD93_RAT STANDARD; PRT; 643 AA.

AC O9ET61; O9J126;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Complement component C1q receptor precursor (Complement component 1, q

DE subcomponent, receptor 1) (C1qR) (C1QR(P)) (C1q/MBL/SPA receptor)

DE (CD93 antigen) (Cell surface antigen AA4).

GN C1QR1 OR CD93 OR C1QR.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PVG; TISSUE=Natural killer cells;

RX MEDLINE=20545218; PubMed=1093152;

RA Lovik G., Vaage J.T., Dissen E., Szpirer C., Ryan J.C., Rolstad B.;

RT "Characterization and molecular cloning of rat C1qR, a receptor on NK

RT cells.";

RL Eur. J. Immunol. 30:3355-3362(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Lung;

RX MEDLINE=20507883; PubMed=10934210;

RA Dean Y.D., McGreal E.P., Akatsu H., Gasque P.;

RT "Molecular and cellular properties of the rat AA4 antigen, a C-type

RT lectin-like receptor with structural homology to thrombomodulin.";

RL J. Biol. Chem. 275:34382-34392(2000).

CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for

CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant

CC protein A (SPA). May mediate the enhancement of phagocytosis in

CC monocytes and macrophages upon interaction with soluble defense

CC collagens. May play a role in intercellular adhesion.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and

CC heart. Expressed at lower level in brain, thymus, liver, spleen,

CC intestine, kidney, adrenal gland, muscle and testis. Expressed on

CC endothelial cells, platelets, undifferentiated monocytes and

CC circulating natural killer cells.

CC -!- PTM: N- and O-glycosylated (By similarity).

CC -!- SIMILARITY: Contains 1 C-type lectin family domain.

CC -!- SIMILARITY: Contains 5 EGF-like domains.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

```

DR EMBL: AF136537; AAG01572.1; -.
DR HSSP: AF160978; AAF80402.1; -.
DR HSSP: P35555; 1EMN.
DR GO: GO:0016021; C: integral to membrane; ISS.
DR GO: GO:0004872; F: receptor activity; ISS.
DR GO: GO:0016337; P: cell-cell adhesion; ISS.
DR GO: GO:0042116; P: macrophage activation; ISS.
DR GO: GO:0006909; P: phagocytosis; ISS.
DR InterPro: IPR000152; ASx_hydroxyl.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00008; EGF; 5.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00179; EGF_CA; 3.
DR PROSITE: PS00010; ASX_HYDROXYL; 3.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR PROSITE: PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01187; EGF_CA; 3.
KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW EGF-like domain; Lectin; Glycoprotein.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 643 COMPLEMENT COMPONENT C1Q RECEPTOR.
FT DOMAIN 24 571 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 572 592 POTENTIAL.
FT DOMAIN 593 643 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 173 C-TYPE LECTIN.
FT DOMAIN 257 298 EGF-LIKE 1.
FT DOMAIN 299 341 EGF-LIKE 2.
FT DOMAIN 342 381 EGF-LIKE 3.
FT DOMAIN 382 423 EGF-LIKE 4.
FT DOMAIN 424 462 EGF-LIKE 5.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 268 282 BY SIMILARITY.
FT DISULFID 284 297 BY SIMILARITY.
FT DISULFID 303 314 BY SIMILARITY.
FT DISULFID 308 325 BY SIMILARITY.
FT DISULFID 327 340 BY SIMILARITY.
FT DISULFID 346 355 BY SIMILARITY.
FT DISULFID 351 364 BY SIMILARITY.
FT DISULFID 366 380 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 393 406 BY SIMILARITY.
FT DISULFID 408 422 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 433 446 BY SIMILARITY.
FT DISULFID 448 461 BY SIMILARITY.
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 417 417 E -> K (IN REF. 2).
SQ SEQUENCE 643 AA; 68781 MW; 9AEAC933AD943DB6 CRC64;

Alignment Scores:
Pred. No.: 0.466 Length: 643
Score: 113.00 Matches: 93
Percent Similarity: 28.12% Conservative: 31
Best Local Similarity: 21.09% Mismatches: 146
Query Match: 2.82% Indels: 171
DB: 1 Gaps: 22

US-09-743-237-1 (1-2241) x CD93_RAT (1-643)
QY . 656 GATTTCAGCTGAAGAGGCGCCAGATGCTCTCATAGACAACTGTGGCGGAGGGA 715
Db ||||||| :||| |||
. 146 AspLeuSerLeu-----LysProHisProSerHisLeuProLysrHisGluSer 162
QY 716 GCTCAAGCGCTCCATCTCTCTCAGTACGATGACGAGCAGTTTCCCTCAGTCAGA 775
Db ||| |||
. 163 ProCysGlyThrProAspAlaProGly-----AsnSerIleGlu 175
QY 776 GTCCTCCTAAGCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTA-----CC 823

```

```

176 GlyPheLeuCysLysPheAsnPheLysGlyMetCysSer-ProLeuAlaLeuGlyGlyPr 195
824 AGCGAGTTAAATCTCATCACACAGGTTGATAATGAGCTCTCCCATCAGCTGCTCAATGG 883
195 oGlyGlnLeuThrTyrThrProPheGln-----AlaThrThrSerSerLeuLysAl 213
884 GGCTGCCTTT----- 893
213 aValProPheAlaSerValAlaAsnValValCysGlyAspGluAlaGluSerLysThrAs 233
894 -----CCCTCTGGAGCTGC 907
233 nTyrTyrLeuCysLysGluThrThrAlaGlyValPheHisTrpGlySerSerGlyPro-- 252
908 TCTGCAAGGCCACCCAAATAACTGTCT-----GGTACTGT-----CACATG 952
253 -LeuCysValSerProLysPheGlyCysSerPheAsnAsnGlyGlyCysGlnGlnAspCy 272
953 CTTCTCCAGCGGGGACTTCTGCAACAGCTGCAACACCTCGCATGAGCTGAGCTCGA 1012
272 sPheGluGlyGlyAspGlySerPheArgCysGlyCys-----ArgProGlyPheAr 289
1013 GCGCTTCAAAGCCATAAAGCGTGTCTTGATAGAAATCTGAAGCTTTCCAAACCAAAAT 1072
289 gLeuLeuAspAspLeuValThrCysAlaSerArgAsnProCysSerSerAsnProCysTh 309
1073 GGGGAAGCGCTCTGGGAGCTGCTAAACTTCGACACACACAAAGGTGCAACTGTAAAGCG 1132
309 rGlyGlyMetCysHisSerValProLeuSerGluAsnTyrThrCysHisCysProAr 329
1133 CTCAGGCTCCCTGAAGAACTACTGTGAGTGTATGAGGCGCAAAATCATGTGTCTTCCAT 1192
329 gGlyTyrGlnLeuAspSer-----SerGlnValHisCysValAspIl 343
1193 TTGCAATCATGTTGCAAAACTATGAAGAACTGAGAACTGCAAGAAATGCTGTATGAG 1252
343 eAspGluCys-----GluAspSerProCysAspGlnGluCysIleAs 357
1253 CACACCCACTACATGAGCGCTGGGACTTTGAGACAGCCATTATTTGTCCCCACCA 1312
357 nThrPro----- 359
1313 GTTCTCAGGAGCTCCAAACTGAGAAATAAGGAGGCTTCTCTGTATCTCTCTG-- 1370
360 -----GlyGlyPheHisCysGluCysrPva 368
1371 -----GAAGTAGTGGAGCCACA-- 1388
368 lGlyTyrGlnSerSerGlySerLysGluAlaCysGluAspValAspGluCysThrAl 388
1389 -----TGTCCCTG 1396
388 aAlaTyrSerProCysAlaGlnGlyCysThrAsnThrAspGlySerPheTyrCysSerCy 408
1397 C-----CTGCTGGCCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1447
408 sLysGluGlyTyrIleMetSerGlyGluAspSerThrGln----- 421
1448 GGCTGAGCAGATGATCTCTGAGGAGTTTGGAGGTCCTGTCGACATTTCTCCACATCGA 1507
422 -----CysGluAspIleAspGluCysLeuGlyAsn----- 431
1508 GTTCAAGTCCAAAGGGGCTGAAAATTCAGTAGCTGCAAGCTGTAAGGGAATGCTGT 1567
432 -----ProCysAspThrLeuCysIleAsnThrAspGlySerPheArgCysGlyCysProAl 450
1568 GGCA---AGCCTCAGCCCTGGGAATCTGCAC-----CGAGGAAGCTGG-----TG 1609
450 aGlyPheGluLeuAlaProAsnGlyValSerCysThrArgGlySerMetPheSerGluLe 470
1610 CCCAGGGAGG-----ACGACAGCGCGGCATCA----- 1637

```

```
Db 470 uProAlaArgProProGlnLysGluAspLysGlyAspGlyLysGluSerThrValProLe 490
Qy 1638 -----TGGCCAGGTGAGCTGTGAGTCTGAGTGTGATGCTGATGCTGAGCCAGCTAC 1690
Db 490 uThrGluMetProGlySerLeuAsnGlySerLysAspValSerAsnArgAlaGlnThrTh 510
Qy 1691 T 1691
Db 510 r 510

RESULT 12
AMFR_HUMAN STANDARD; PRT; 323 AA.
AC P26442;
AT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Autocrine motility factor receptor precursor (AMF receptor) (GP78).
GN AMFR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91302382; PubMed=1649192;
RA Watanabe H., Carmi P., Hogan V., Raz T., Silletti S., Nabi I.R.,
RA Raz A.;
RT "Purification of human tumor cell autocrine motility factor and
RT molecular cloning of its receptor.";
RL J. Biol. Chem. 266:13442-13448(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95352090; PubMed=7626106;
RA Huang B., Xie Y., Raz A.;
RT "Identification of an upstream region that controls the transcription
RT of the human autocrine motility factor receptor.";
RL Biochem. Biophys. Res. Commun. 212:727-742(1995).
CC -!- FUNCTION: SPECIFIC RECEPTOR FOR THE AUTOCRINE MOTILITY FACTOR.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: PHOSPHORYLATED IN THE PRESENCE OF AMF.
CC -!- PTM: O-GLYCOSYLATED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M63175; AAA36671.1; -
DR EMBL; L35233; AAA79362.1; -
DR PIR; A39877; A39877.
DR Genew; HGNC:463; AMFR.
DR MIN; G03243; -
DR GO; GO:0006928; P:cell motility; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
KW Receptor; Glycoprotein; Phosphorylation; Signal; Transmembrane;
KW Polymorphism.
KW SIGNAL 1 17
FT CHAIN 18 323
FT DOMAIN 18 110
FT TRANSF 111 137
FT TRANSF 138 323
FT CYTOPLAS 138 323
FT N-LINKED 138 323
FT CARBOHYD 24 194
FT MOD_RES 194 194
FT VARIANT 181 181
FT I -> V (IN dbSNP:4924).
FT /FTID=VAR_014809.
SQ SEQUENCE 323 AA; 34325 MW; 0A7AF4DCF90A8700 CRC64;
```

Alignment Scores:

```
Pred. No.: 0.426 Length: 323
Score: 112.50 Matches: 86
Percent Similarity: 36.68% Conservative: 42
Best Local Similarity: 24.64% Mismatches: 122
Query Match: 2.81% Indels: 99
DB: 1 Gaps: 18

US-09-743-237-1 (1-2241) x AMFR_HUMAN (1-323)
Qy 444 TCCGATGAACCTACTGCGCCAGGTTCCTCCAGGACGAGCAAGCAAGCAAGAAAGTGTGAAATC 503
Db 33 SerGluSerPheLeuPro---SerGluGlyAlaSerSerAspProValThrLeuArgArg 51
Qy 504 AAAGAAGCAGGTGGTGTGAGTGTCCAGCGCGGAGCCCTGAAAGACGAGCTTTCCAGCGCCCT 563
Db 52 ArgMetLeuAlaAlaAlaArgAsnGlyGlyPheArgSerSerArgProProSerAlaPro 71
Qy 564 CTGGCTCAGGAATCTGTTGCAAGTTCCTCCATCATCCAGGAGGAGGAGGAGGAGGAGGAGGAG 623
Db 72 Leu-----ProSerSer-----AlaAlaSerCysAlaLeu 81
Qy 624 TGCCCT-----CGGAAGAAAGACTCCAGAGCCCATGGTGTGATTGTGCAGCTGAAAGGA 674
Db 82 CysProThrAspTrpArgArg-----ProValProLeuLeuProLeuHisGly 97
Qy 675 GGCCGCCAGATGCTCTGCATAGACAACATGTGGCGCGAGGAGGAGCTCAAAGCGCTCCATCTG 734
Db 98 LysAla-----GlyLeuThrAla 103
Qy 735 CTTCCTCAGTACGATGACGAGCAGTTCCTCCTCAGTACAGAGCTCCCTTAAGCCAAATG--- 791
Db 104 LeuProLeuTyrLysAlaCysGlyLeuValPheGlyGlnLeuLeuAsnLeuLeu 123
Qy 792 -----ACAACTTTAGTGGGAAGACTTCTGCGCAGTACCAGCGAAGTTAAATCTCATC 842
Db 124 LeuCysAsnThrPheTyrValThrPheLeuPheProLeuGluThr---LeuGlnIleLeu 142
Qy 843 ACACAGGTGATAATGAGCTCTCCCATCAGCTCTCAATGCGGTGCTTCCCTCTCGGA 902
Db 143 ThrVal-----GlyMetIleSerSerGlyValAspTrpThrAlaTrpGlyGly 159
Qy 903 CCTGCTCTGCNAGGGCCACCCAAATACTCTGCTGGTACTGTGCTGCTTCTTCCA-- 960
Db 160 ArgSer---GlyGlySerGluProValAla--CysLeuGlnGlnAlaAlaSerThrProAl 178
Qy 961 -----GCGGGACTCTCTGCAACAGCTGCAGCTGCAGTCAACAACTG 998
Db 178 aSerCysIleArgProThrAsnAlaGlyValLeuSerThrThrProSerGlyLysSerVa 198
Qy 999 CGCCATGAGCTCGAGCGCTTCAAAGCCATAAGCGGTCTCTTGTATAGAAATCTCGAAGCT 1058
Db 198 lGlyGluAlaHisSerValSerProProArgGlyValThrSerValIleLysLe 218
Qy 1059 TTCCAACCAAAATGGGAA-----AGCCGCTCTGGAGCTGCT 1097
Db 218 uLeuSerLeuLeuTrp-LysHisValAspCysAlaArgAlaArgProThrGlySerCys- 237
Qy 1098 AAAGTTCGACACAGCAAGGTTGCAACTGTAAAGCGCTCAGGCTGCCTGCCTGAAGAACTACTGT 1157
Db 238 --ThrProGluGlnGlnGlyLe-----LeuGluLysGluLeuLeu- 250
Qy 1158 GAGTGTATGAGGCCCAAAATCATCTGTTCTTCCATTTTGCAAATGCATGCTGTGCAAAAC 1217
Db 251 -----ValArgTyrLeuGluGlnA 257
Qy 1218 TATCAAGAAATCCAGAACGAAAAATGCTGTATGAGCACACCCACTACATGAGCCCTGGG 1277
Db 257 rgArgGlyLysSerArgAlaIleGlyCysaspGluValThrPro----- 271
Qy 1278 GACTTTGAGACAGCCCATTTATTTGTCGCCACCAAGTTCTCAGGACCTCAAAACTGAG- 1336
Db 272 -----PheCysProThr-----ThrSerGlyThrAsp 281
```


DR	ENBL; X06061; CAA29454.1; JOINED.
DR	ENBL; X06062; CAA29454.1; JOINED.
DR	ENBL; X06063; CAA29454.1; JOINED.
DR	ENBL; X06064; CAA29454.1; JOINED.
DR	ENBL; X06065; CAA29454.1; JOINED.
DR	ENBL; X06066; CAA29454.1; JOINED.
DR	ENBL; X06067; CAA29455.1; -
DR	ENBL; X06068; CAA29455.1; JOINED.
DR	ENBL; X06069; CAA29456.1; -
DR	ENBL; X06070; CAA29456.1; JOINED.
DR	ENBL; X02749; CAA26527.1; -
DR	ENBL; AF170489; RAD51647.1; -
DR	ENBL; AF170486; RAD51647.1; JOINED.
DR	ENBL; AF170487; RAD51647.1; JOINED.
DR	ENBL; AF170488; RAD51647.1; JOINED.
DR	ENBL; AF105687; AAC95473.1; -
DR	ENBL; AF105681; AAC95473.1; JOINED.
DR	ENBL; AF105682; AAC95473.1; JOINED.
DR	ENBL; AF105683; AAC95473.1; JOINED.
DR	ENBL; AF105684; AAC95473.1; JOINED.
DR	ENBL; AF105685; AAC95473.1; JOINED.
DR	ENBL; AF105686; AAC95473.1; JOINED.
DR	ENBL; AF080484; RAD50912.2; -
DR	ENBL; AF169654; RAD50912.2; JOINED.
DR	ENBL; AF169655; RAD50912.2; JOINED.
DR	ENBL; AF169656; RAD50912.2; JOINED.
DR	ENBL; AF169657; RAD50912.2; JOINED.
DR	ENBL; AF169658; RAD50912.2; JOINED.
DR	ENBL; AF169659; RAD50912.2; JOINED.
DR	ENBL; AF169661; RAD50912.2; JOINED.
DR	ENBL; AF169662; RAD50912.2; JOINED.
DR	ENBL; AF169663; RAD50912.2; JOINED.
DR	ENBL; AF169664; RAD50912.2; JOINED.
DR	ENBL; AF080472; RAD50912.2; JOINED.
DR	ENBL; AF080473; RAD50912.2; JOINED.
DR	ENBL; AF080474; RAD50912.2; JOINED.
DR	ENBL; AF080475; RAD50912.2; JOINED.
DR	ENBL; AF080476; RAD50912.2; JOINED.
DR	ENBL; AF080477; RAD50912.2; JOINED.
DR	ENBL; AF080478; RAD50912.2; JOINED.
DR	ENBL; AF080479; RAD50912.2; JOINED.
DR	ENBL; AF080480; RAD50912.2; JOINED.
DR	ENBL; AF080481; RAD50912.2; JOINED.
DR	ENBL; AF080482; RAD50912.2; JOINED.
DR	ENBL; AF080483; RAD50912.2; JOINED.
DR	PIR; A59110; UIHU.
DR	HSSP; P21836; lMAA.
DR	Genew; HGNC:11764; TG.
DR	MTM; 188450; -
DR	MTM; 138800; -
DR	MTM; 274900; -
DR	InterPro; IPR002018; CarbesteraseB.
DR	InterPro; IPR007016; Thyroglobulin_1.
DR	Pfam; PF00135; Coesterase; 1.
DR	Pfam; PF00086; thyroglobulin_1; 9.
DR	SMART; SM00211; TY; 10
DR	PROSITE; PS00484; THYROGLOBULIN_1; 9.
DR	PROSITE; P500941; CARBOXYLESTERASE_B_2; 1.
DR	KW Glycoprotein; Repeat; Thyroid hormone; Iodination; Susceptibility to disease mutation; Polymorphism.
FT	SIGNAL 1 19
FT	CHAIN 20 2768
FT	DOMAIN 32 92
FT	DOMAIN 93 160
FT	THYROGLOBULIN TYPE IA 1.1
FT	THYROGLOBULIN TYPE IA 2.1
Alignment Scores:	
Pred. No.:	0.752
Score:	Length: 2768
	Matches: 143
Percent Similarity:	Conservative: 82
Best Local Similarity:	Mismatches: 270
Query Match:	Indels: 195
DB:	Gaps: 35

QY 1233 GAACGAAATGCTGATGACACACCC-----CACTACATGAGCGCTGGGACTTT 1283
Db 1486 GlySerLeuAlaCysValProCysProValGlyArgThrThrIleSerAlaGlyAlaPhe 1505
QY 1284 GAGACGACCATATTATTGTCCCA----- 1307
Db 1506 SerGlnThrHisCysValThrAspCysGlnArgAsnGluAlaGlyLeuGlnCysAspGln 1525
QY 1308 ---GCCAAGTCTCAGGACCTCCAAACTGAGAAAATAGCAGCGCTTCCTCCTGATC 1364
Db 1526 AsnGlyGlnThrArgAlaSerGlnLysAspArgGlySerGlyLysAlaPhe---CysVal 1544
QY 1365 TCC-----TGGGAAGTAGTGGAGGCC-----ACATGTGCC 1394
Db 1545 AspGlyGluGlyArgArgLeuProThrPrpGluThrGluAlaProLeuGluAspSerGln 1564
QY 1395 TGCCTGCTGGCCCGGCTGAGAA----- 1418
Db 1565 CysLeuMetMetGlnLysPheGluLysValProGluSerLysValIlePheAspAlaAsn 1584
QY 1419 -----GCAGACGAGGACACGTGTTCCCAAGCTTGGCTGACGATGATCCTGGAG 1469
Db 1585 AlaProValAla-ValArgSerLysValProAspSerGluPhePro-----ValMetG1 1602
QY 1470 GAGTTTG-----GAAGGTGCTGTCGCAGATTCTCCACATCAGTTCAG 1514
Db 1602 nCysLeuThrAspCysThrGluAspGluAlaCysSerPhePheThrValSerThrG1 1622
QY 1515 TCCAGGGGCTGAAATGAGTAGGTGCTGCAAGCTGGTAAGGGGAATGC-----CT 1565
Db 1622 uPro-----GluIleSerCysAspPheThrAl 1631
QY 1566 GTGCAAGCCTCAGCCCTGGGAATCTGCACC-----GAGGAAGCTGTGCTGCCAGGAGG 1619
Db 1631 aThrPThrSerAspAsnValAlaCysMetThrSerAspGlnLysArgAspAlaLeuGlyAs 1651
QY 1620 AGCAGGCGCGCATCATGTCAGGTGCTGAGGCTGCTGAGTAT-CTGCATGGTAC 1678
Db 1651 nSerLysAlaThrSerPheGlySerLeuArgCysGlnValLysValArgSerHisGlyG1 1671
QY 1679 TGGCCAGCCTACT-----CAAGGTATCTTAAAGTCAAGCAGCAGCAGA 1720
Db 1671 nAspSerProAlaValTyrLeuLysLysGlyGlnGlySerThrThrThrLeuGlnLysar 1691
QY 1721 G---CCACCTCGGGATGACACTGGCCCTCTGCTGCCCTGGGAGGCC----- 1766
Db 1691 gPheGluProThrGlyPheGlnAsnMetLeuSerGlyLeuTyrAsnProIleValPheSe 1711
QY 1767 ----TCTGGGACTCCCTGCC-----TGCAATAAAGAGGCTGATTTTC 1807
Db 1711 rAlaSerGlyAlaAsnLeuThrAspAlaHisLeuPheCysLeuLeuAlaCysAspArgAs 1731
QY 1808 TACTTGTGTTGTTGTTGCTTTCAAATGCTTAGTAGTACTCCTCAATCAAGTT---AT 1864
Db 1731 pLeuCysCysAspGlyPheValLeuThrGlnValGlnGlyAlaIleCysGlyLe 1751
QY 1865 TATGACCCAGCCTCAAGTTAGAGACTAGGCTCTTCTCAGGTGGACTCTGCCCAATCA 1924
Db 1751 uLeuSerSerProSerValLeu---LeuCysAsnValLysAspTrpMetAspProSerG1 1770
QY 1925 CATACAAGTCAGGTGCCATCAGGCTTTTCCAGGCCAGGCTGTCGACAGGATATGG 1984
Db 1770 u-----AlaTrpAlaAsnAlaThrCysProGlyValThrTyrAspGlnGlu---Se 1786
QY 1985 GAGGGGGTGGGTTAGAGCTGGT 2009
Db 1786 rHisGlnValIleLeuArgLeuGly 1794
RESULT 14
CD93_HUMAN
ID CD93_HUMAN
AC Q9NPY3; O00274; PRT; 652 AA.

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement component C1q receptor precursor (Complement component 1, q
DE subcomponent, receptor 1) (C1qR) (C1qR(P)) (C1q/MBL/SPA receptor)
DE (CD93 antigen) (CDw93).
GN C1QR1 OR CD93.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA Nepomuceno R.R., PubMed=9047234;
RT "cDNA cloning and primary structure analysis of C1qR(P), the human
RT C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro";
RL Immunity 6:119-129(1997).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT ALA-318.
RX MEDLINE=21640567; PubMed=11781389;
RA Steinberger P., Szekeres A., Wille S., Stockl J., Selenko N.,
RA Prager E., Staffler G., Madic O., Stockinger H., Knapp W.;
RT "Identification of human CD93 as the phagocytic C1q receptor (C1qR)
RT by expression cloning";
RL J. Leukoc. Biol. 71:133-140(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehaeslaitho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20";
RL Nature 414:865-871(2001).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan H., Farmer A., Rubin G.M., Hong L.,
RA Diatchenko L., Marusina K., Bontal M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallick D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.:
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=21990337; PubMed=11994479;
RA McGreal E.P., Ikegaki N., Akatsu H., Morgan B.P., Gasque P.:
RT "Human Clqrp is identical with CD93 and the mni-11 antigen but does
RT not bind Clq.";
RL J. Immunol. 168:5222-5232(2002).
RN [6]
RP O-GLYCOSYLATION.
RX MEDLINE=99192777; PubMed=10092817;
RA Nepomuceno R.R., Ruiz S., Park M., Tenner A.J.:
RT "Clqrp is a heavily O-glycosylated cell surface protein involved in
RT the regulation of phagocytic activity.";
RL J. Immunol. 162:3583-3589(1999).
CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
CC Clq, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SPA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Highly expressed in endothelial cells,
CC platelets, cells of myeloid origin, such as monocytes and
CC neutrophils. Not expressed in cells of lymphoid origin.
CC -!- PTM: N- and O-glycosylated.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 5 EGF-like domains.
CC -!- CAUTION: Has been sometimes referred to as a collectin receptor.
CC -!- CAUTION: According to Ref.5, Clq is not a ligand for ClqR1.
CC -!- DATABASE: NAME=PROW; NOTE=PROW 3.1-6(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/467246456.g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U94333; AAB53110.1; -;
DR EMBL; AL118508; CAC00597.1; -;
DR EMBL; BC028075; AAH28075.1; -;
DR HSSP; P35555; 1EMN.
DR Genew; HGNC:15855; ClqR1.
DR MIM; 120577; -;
DR GO; GO:0016021; C:integral to membrane; IC.
DR GO; GO:0004872; F:receptor activity; NAS.
DR GO; GO:0016337; P:cell-cell adhesion; IDA.
DR GO; GO:0042116; P:macrophage activation; NAS.
DR GO; GO:0006909; P:phagocytosis; NAS.
DR InterPro; IPR000152; Asx_hydroxy1.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00008; Lectin_5.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF01108; Tissue_fac; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW EGF-like domain; Lectin; Glycoprotein; Polymorphism.

FT	SIGNAL	1	21		
FT	CHAIN	22	652		COMPLEMENT COMPONENT C1Q RECEPTOR.
FT	DOMAIN	24	580		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	581	601		POTENTIAL.
FT	DOMAIN	602	652		CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	32	174		C-TYPE LECTIN.
FT	DOMAIN	260	301		EGF-LIKE 1.
FT	DOMAIN	302	344		EGF-LIKE 2.
FT	DOMAIN	345	384		EGF-LIKE 3.
FT	DOMAIN	385	426		EGF-LIKE 4.
FT	DOMAIN	427	468		EGF-LIKE 5.
FT	DOMAIN	594	601		POLY-LEU.
FT	DISULFID	264	275		BY SIMILARITY.
FT	DISULFID	271	285		BY SIMILARITY.
FT	DISULFID	287	300		BY SIMILARITY.
FT	DISULFID	306	317		BY SIMILARITY.
FT	DISULFID	311	328		BY SIMILARITY.
FT	DISULFID	330	343		BY SIMILARITY.
FT	DISULFID	349	358		BY SIMILARITY.
FT	DISULFID	354	367		BY SIMILARITY.
FT	DISULFID	369	383		BY SIMILARITY.
FT	DISULFID	389	400		BY SIMILARITY.
FT	DISULFID	396	409		BY SIMILARITY.
FT	DISULFID	411	425		BY SIMILARITY.
FT	DISULFID	431	443		BY SIMILARITY.
FT	DISULFID	439	452		BY SIMILARITY.
FT	DISULFID	454	467		BY SIMILARITY.
FT	CARBOHYD	325	325		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	318	318		V -> A.
FT	CONFLICT	22	22		/FTID=VAR_013573.
FT	CONFLICT	36	36		T -> V (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	38	38		C -> T (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	155	155		TA -> RI (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	186	186		S -> N (IN REF. 1).
FT	CONFLICT	186	186		G -> A (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	492	492		S -> A (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	496	496		R -> Q (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	504	504		R -> G (IN REF. 1; AA SEQUENCE).
FT	CONFLICT	541	541		P -> S (IN REF. 1).
SQ	SEQUENCE	652 AA;	68560 MW;		BECA0FEAC55FCAC2 CRC64;

Alignment Scores:
Pred. No.: 0.562 Length: 652
Score: 112.00 Matches: 129
Percent Similarity: 27.11% Conservative: 54
Best Local Similarity: 19.11% Mismatches: 183
Query Match: 2.79% Indels: 309
DB: 1 Gaps: 30

US-09-743-237-1 (1-2241) x CD93_HUMAN (1-652)

Qy	39	GGCTCAGGCACACACAGGATACACAG	-----TGTGTGTTCTCGCCTGTGGACT	89
Db	19	GlyAlaGlyThrGlyAlaAspThrGluAlaValValCysValGlyThrAlaCysTyrThr	38	
Qy	90	TGTGACTCCACCCACCTCCGCCCCAGCAGGCT	-----	122
Db	39	AlaHisSerGlyLysLeuSerAlaAlaGluAlaGlnAsnHisCysAsnGlnAsnGlygly	58	
Qy	123	-----AGGATAGAACCCAGGCCCTTTTCGTTGCTCG	---CAG	158
Db	59	AsnLeuAlaThrValLysSerLysGluAlaGlnHisValGlnArgValLeuAlaGln	78	
Qy	159	ATAGTC-----	-----	164
Db	79	LeuLeuArgArgGluAlaAlaLeuThrAlaArgMetSerLysPheTrpIleGlyLeuGln	98	
Qy	165	-----TTACGCTGGTAGTTTGGGTTGCTGGGA	194	
Db	99	ArgGluLysGlyLysCysLeuAspSerLeuProLeuLysGlyPheSerTrpValGly	118	
Qy	195	-----GATTTTTTTTCTTCACACCAAGACTTCCATTATTAGAGATTTTTCAGT	245	


```
Db 119 GlyGlyGluAsp-----ThrProTyrSerAsnTrpHisLysGlu----- 131
QY 246 TGATGATCTCCCCCTCTGTAAGATAAGGACAGTTCTTTAAACCTATGTAGAGTTTGA 305
Db 131 ----- 131
QY 306 TGAATTCCTGCTTCAACATATTCCTAAGCTATATAGCAATTCCTTGAATTCGTATAT 365
Db 132 -----LeuArgAsnSerCysIleSerLys-----ArgCysValSerLeuLeuLeu 146
QY 366 RACTTAGGAGAACCTCTGATCTCTGCTCTACATCTCTAGCTAGCTAGCTAGCTAGCT 425
Db 147 AspLeuSerGlnProLeuLeuProSerArgLeu-----ProLys----- 159
QY 426 AAATCATTTTGGTGAGACTCCGATCAACTACTGCGAGGTTCCTCAAGCAGCAGCAAGCA 485
Db 160 -----TrpSerGluGlyProCysGlySerProGlySerProGlySer----- 173
QY 486 AGAAAAGTGTGAATCAAGAACAGCTGGTGTAGTGTGCCAGGC----- 530
Db 174 -----AsnIleGluGlyPheValCysLysPheSerPheLysGlyMetCysArgProLeu 191
QY 531 -----GGCAGCCCTGAAGACCCAGCTTTCCAGGCCCTCTGGCTCAGGAATCTCTGTC 584
Db 192 AlaLeuGlyGlyProGlyGlnValThrTyrThrProPheGlnThrThrSer----- 209
QY 585 AAGTTCCCATCATCCAGGAGCAGAGGAGGCTCCAGCTGC----- 626
Db 210 -----SerSerLeuGluAlaValProPheAlaSerAlaAlaAsnValAlaCysGly 226
QY 627 -----CCTCGGAAGAACTCCAGCCCATCGTGGTATTTGTCAGCTGAAA----- 671
Db 227 GluGlyAspLysAspGluThrGlnSerHisTyrPheLeuCysLysGluLysAlaProAsp 246
QY 672 -----GGAGGCGCCAGATGCTCTCCATAGACAACCTGTGGCGCAGGAGCTC 719
Db 247 ValPheAspTrpGlySerSerGlyProLeuCysValSer----- 259
QY 720 AAAGCGCTCCATCTGCTCTCCATGATGATGACAGAGCAGTTTCCCTCAGTCAGAGCTC 779
Db 260 -----ProLysTyr----- 262
QY 780 CCTAAGCCAAATGACAACCTTTAGTGGGAAGACTTTCGCCAGTACACAGGAAGTTAAATCTC 839
Db 262 ----- 262
QY 840 ATCACACAGTTGATAATGAGCTCTCCCATCAGCTGTCAATGGGGCTGCCCTTCCCTCT 899
Db 263 -----GlyCysAsnPheAsnAsn 268
QY 900 GGACCTGCTCTGCAAGGGCCACCCAAATAACTCTGTCTGGTACTGTGACTGCTCTCC 959
Db 269 GlyGlyCysHisGln-----AspCysPheGlu 277
QY 960 AGCGGGGACTTCTGCAACAGCTGCAAGCTGCAACAACCTGCGGCATGAGCTCAGAGCGCTTC 1019
Db 278 GlyGlyAspGlySerPheLeuCysGlyCys-----ArgProGlyPheArgLeuLeu 294
QY 1020 AAAGCCATAAAGCGCTGCTTCTGATAGAAATCTGAAGCTTTCCACCAAAAATGGGAAA 1079
Db 295 AspAspLeuValThrCysAlaSerArgAsnProCysSerSerProCysArgGly--- 313
QY 1080 GGCCGCTCTGGAGCTGCTAAACTTGCACACAAAGGCTGCAACTGTAAAGCGCTCAGGC 1139
Db 314 -----GlyAlaThrCysValLeuGlyPro 321
QY 1140 TGCCTGAGAACTAC---TGTGAGTGCATATAGGCCAAATCATGTTCTTCCATTTGC 1196
Db 322 HisGlyLysAsnTyrThrCysArgCysProGlnGlyTyrGlnLeuAspSerSerGlnLeu 341
QY 1197 AAATGCATTCGTCGAAAACCTATGAGNAAGTCCAGACGAAATAATGCTGATGAGCACA 1256
Db 342 AspCysValAspValAspGluCysGlnAspSerProCysAlaGlnGluCysValAsnThr 361
```

```
QY 1257 CCCACTACATGGAGCGCTGGGACTTTGAGAGCAGCCATTATTTGCCCCAGCAAGTTC 1316
Db 362 Pro----- 362
QY 1317 TCAGGACCTCCAAAACCTAGAAAAAATAGGCAAGGCTTCTCTCTGTATCTCTCTGGGAAGTA 1376
Db 363 -----GlyGlyPheArgCysGluCysTrpValGly 372
QY 1377 GTGGAGGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1436
Db 373 TyrGlu-ProGlyGlyProGly----- 379
QY 1437 TCCCAAGCTTGGCTGAGCAGATGATCTCTGGAGGAGTTTGAAGTGGCTGCTGCTGCTGCT 1496
Db 380 -----GluGlyAlaCysGln----- 384
QY 1497 CTCACATCGAGTTCAAGTCCAAAGGGCTGAAAAATAGTAGCTAGCTGCAAGCTGGTAAAGG 1556
Db 385 -----AspValAspGluCysAlaLeuGlyArgSe 394
QY 1557 GGAATGCTGTGGCAAGCTCTAGCCCTGGGAATCTGCACCGAGGAAGCTGGTGCC----- 1611
Db 394 rProCys-----AlaGlnGly---CysThrAsnThrAspGlySerPheH 408
QY 1612 -----CAGGAGGAGCAGAGCGCCGCATCATGCCAGGTGAGTGTGAGT 1658
Db 408 scysSerCysGluGluGlyTyrValLeuAlaGlyGluAspGlyThr--GlnCysGlnAsp 427
QY 1659 CTGAGTGTATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1694
Db 428 ValAspGluCysValGlyProGlyGlyProLeuCysAspSerLeuCysPheAsnThrGln 447
QY 1695 GGTATCTTAAGTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1733
Db 448 GlySerPheHisCys---GlyCysLeuProGlyTrpValLeuAlaProAsnGlyValSer 466
QY 1734 TGGACACGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1772
Db 467 CysThrMetGlyProValSerLeuGlyProProSerGly 479
RESULT 15
IP3L_HUMAN
ID IP3L_HUMAN STANDARD; PRT: 946 AA.
AC P27987; O96J51; O9UH47;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Inositol-trisphosphate 3-kinase B (EC 2.7.1.127) (Inositol 1,4,5-
DE trisphosphate 3-kinase) (IP3K) (IP3 3-kinase) (IP3K-B).
GN ITPKB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RC TISSUE=Hippocampus;
RC SEQUENCE FROM N.A.
RA Dewaste V.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RT "Cloning of the complete protein coding regions for inositol 1,4,5-
RT trisphosphate 3-kinase B-isoforms from rat and human.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 257-946 FROM N.A.
RA Donnelly S.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 442-946 FROM N.A.
```



```

QY 1748 CCTGTCCCTGGGAGGCCCTCTGGG-----GACTCCCTGCCCTGCATAAAAAGAGG 1798
Db   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
465 nValGluAlaGlyIleProSerGlyArgMetLeuGluProLeuProCys----- 481
QY 1799 GTGATTTTCTACTTGTGTGTATGTGTTGCTTTCAAATTGCTTAGTAGTACCTCCATTCA 1858
Db -----TTPAspAlaAl 485
QY 1859 AGTTATTATGACCCAGCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGGACTCTGCC 1918
Db   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
485 aLysAspLeuLysGluProGln-----CysPr 494
QY 1919 AAATCACATACAAAGTCAGGTGCCATCAGGGGTTTTTCCAGGCCAGGCTGTGCACAGAG 1978
Db   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
494 oProGlyAspArgVal-----GlyValGlnProGlyAsnSer-----Arg-- 507
QY 1979 ATATGGAGGGGGTCCGGGTTAGAGCTGGGTTGTTGG 2017
Db   ||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
508 valTrpGlnGlyThrMetGluLysAlaGlyLeuAlaTrp 520

```

Search completed: July 24, 2003, 14:17:44
Job time : 75.5999 secs


```
QY 1059 TTCAACCAAAATGGGAAGCCGCTCGGAGCTGCTAAA----- 1100
    |||
Db 175 PheArgProLysIleAlaAlaSerProHisGlyArgAspAsnArgGluGluValGly 194
    |||
QY 1101 -----CTTGACACACGAAGGTCAGCTGTAAGCGCTCAGCGCTCGCTG 1145
    |||
Db 195 AspValValMetLeuAlaArgHisAsnLysGlyCysHisCysLysSerGlyCysLeu 214
    |||
QY 1146 AACAACCTACTGTCAGTGCTATGAGGCCAAATCATGCTGCTTCCATTGCAATGCNTT 1205
    |||
Db 215 LysLysTyrCysGluCysPheGlnAlaAsnIleLeuCysSerGluAsnCysLysCysLeu 234
    |||
QY 1206 GCTTCAAAAATATGAGAAATCCAGAACGAAAAATGCTGATGAGCACACCCAC--- 1262
    |||
Db 235 AspCysLysAsnPheGluGlySerGluValArgGlnSerLeuPheHisGlyGluHisSer 254
    |||
QY 1263 -----TACATGAG-----CCTGGGAGCTTTGAGAGCAGC 1292
    |||
Db 255 HisAsnLeuAlaTyrLeuGlnHisAlaAsnAlaIleThrGlyAlaIleGlySerSer 274
    |||
QY 1293 CATTATTTCGCCAGCCAGTTCTTCAGGACCTCCAAAATCTGAGAAAAATAGCAGGCC 1352
    |||
Db 275 GlyPheAlaSer-----AlaProProLysArgArgLysGlyGlnGluIle 290
    |||
QY 1353 TTC 1355
    |||
Db 291 Phe 291

RESULT 3
T08955
hypothetical protein F19B15.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08955
Submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16519
A:Accession: T08955
A:Molecule type: DNA
A:Residues: 1-603 <BEV>
A:Cross-references: EMBL:AL078470; GSPDB:GNO0062; ATSP:F19B15.30
A:Experimental source: cultivar Columbia; BAC clone F19B15
C:Genetics:
Query Match: 9.13%
A:Gene: ATSP:F19B15.30
A:Map position: 4
A:Introns: 107/2; 144/2; 203/3; 321/3; 356/2; 388/1; 522/1

Alignment Scores:
Pred. No.: 8,06e-22 Length: 603
Score: 366.00 Matches: 99
Percent Similarity: 47.56% Conservative: 47
Best Local Similarity: 32.25% Mismatches: 97
Query Match: 9.13% Indels: 64
DB: 2 Gaps: 15

US-09-743-237-1 (1-2241) x T08955 (1-603)
QY 585 AAGTTCCTCATCCAGGAGGCGAGGAGGCGCTCCAGCTGCGCTCGGAAGAAGACTCC 644
    |||
Db 9 GluPheProLysLysAspGlyValGlu---GluGlyPheProThrLysLysProAla 27
    |||
QY 645 AGCCCATGCTGATTTGTCTGCAAGGAGGCGCCAGATGCTCTGCATAGACAACCTGT 704
    |||
Db 28 ArgGlnLeu-----AspPheThrGlyGlySerAspGluHisSerLeuSerLysPro 44
    |||
QY 705 GGCGCGGAGGAGCTCAAGCGCTCCATCTGCTTCCTCAGTACGATGACCAGACGCTTC 764.
    |||
Db 45 AlaAlaProThrValValAlaThrSerValLysProIleIle-----SerSerSerVal 62
    |||
QY 765 CCTCACTCAGAGCTCCTTAGCCCAATCAGCACTTTAGTGGGAAGACTTCTGCCAGTACCA 824
    |||
Db 63 ProSerThrIleArgPro---GlyMetThrIleAlaIleGlyGlnValThrGlnValArg 81
    |||
```

```
QY 825 GCAGAGTTAAATCTCATCACACAGGTTGATAATGAGGCTCTCCCATCAGCTGTCATGGG 884
    |||
Db 82 ProThrLeuProMetAlaThrThrMetSerAsnProProSerGlnSerGlnIleValAsn 101
    |||
QY 885 GCT-----GCCTTTCCTCTGGACCTGCTCTGCAAGGGCCA---CCCAMAATA 929
    |||
Db 102 AlaProIleArgHisProIleProGluSerProLysAlaArgGlyProAlaGProAsnVal 121
    |||
QY 930 -----ACT 932
    |||
Db 122 GluGlyArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHisSerArgCys 141
    |||
QY 933 CTCTCTGGGTACTGTCGACTGCTTCCAGCGGAGCTTCTCCAACACGCTCAGCTGC--- 989
    |||
Db 142 LeuLysLeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCysAsnCysVal 161
    |||
QY 990 -----AACAACTGCGCCATGAGCTCGACGCTTCAAGGCCATAAAGGCGTGTCTT 1040
    |||
Db 162 AsnCysPheAsnAsnValAspAsnGluProAlaArgArgGluAlaValGluAlaThrLeu 181
    |||
QY 1041 GATAGAAATCCTCAAGCTTTCACCAACCAAAATGGGAAA-----GGCCGT--- 1085
    |||
Db 182 GluArgAsnProPheAlaPheArgProLysIleAlaSerSerProHisGlyGlyArgAsp 201
    |||
QY 1086 -----CTGGGAGCTGCTAAACTT-----CGACACAGCAAGGCTCAACTGT 1127
    |||
Db 202 LysArgGluAspIleGlyGluValValLeuLeuGlyLysHisAsnLysGlyCysHisCys 221
    |||
QY 1128 AAGCGCTCAGGCTGCCTGAAGAACTACTGTGAGTGTCTATGAGGCCAAATCATGTGTCT 1187
    |||
Db 222 LysLysSerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIleLeuCysSer 241
    |||
QY 1188 TCATTGTCAAATGCATTGCTTGCAGAACTATGAAGAACTCCAGAACCAAAATGCTG 1247
    |||
Db 242 GluAsnCysLysCysLeuAspCysLysAsnPheGluGlySerGluGluArgGlnAlaLeu 261
    |||
QY 1248 ATGAGCACACCCAC-----TACATGGAG----- 1271
    |||
Db 262 PheHisGlyGluHisSerAsnHisMetAlaTyrLeuGlnGlnAlaAlaAsnAlaIle 281
    |||
QY 1272 CTGCGGAGCTTTGAGAGCAGCCATTTATTTG---TCCCCAGCCCAAGTTCTCAGGACCTCCA 1328
    |||
Db 282 ThrGlyAlaValGlySerSerGlyPheAlaProSerProAla-----Pro 296
    |||
QY 1329 ARACTGAGAAAAATAGGCAG 1349
    |||
Db 297 LysArgArgLysGlyGlnGlu 303
    |||

RESULT 4
F71410
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998
C:Accession: F71410
#sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D.
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
erhoft, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: F71410
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-658 <BEV>
A:Cross-references: GB:297337; NID:g2244829; PID:e326824; PID:g2244834
C:Genetics:
A:Map position: 4COP9-4G3845
```


Db 455 CysGluProGlyValAsnValAsnCysThrAlaGlyThrCysArgLeuValGluAspCys 474
 QY 1005 GAGCTCGAGCGCTTCAAGCCATAAAGCGTCTCTGTATAGAAATCCCTGAGCTTTCCAA 1064
 Db 475 ArgIleArgGlyCysProAlaValProLysCysIleAspArgAspProLeuProPro 494
 QY 1065 CCA-----AAATGGGAAAGCGCTCTGGAGCTGTAACTTCGACACAGC 1112
 Db 495 ProProAspValCysProValGlyThrProValLeuGlyAla-----AspLeu 510
 QY 1113 AAAGGTCGACTGTAAAGCGCTCAGC-----TCCTGAAGAAC---TACTGTGAGTGC 1163
 Db 511 LysGlnLeuTyrCysGlyArgGlyArgCysProTyrPasnThrTyrCysValIle 530
 QY 1164 TATGAGGC-----1172
 Db 531 HisProAlaAspArgTyrAlaValCysCysPheGlySerGlyProAlaSerAlaIleAla 550
 QY 1173 -----AAATCATGTGTCT 1187
 Db 551 ProThrSerAlaProGlyProValAspProCysGluProGlyValAsnValAsnCysThr 570
 QY 1188 ---TCCATTGCAATGATCTGTCGAAACTATGAAGAAGTCCAGAACGAAATG 1244
 Db 571 IleGlyValCysArgLeuValAlaAsnCysAspTyrTrpProCysProAlaArgProThr 590
 QY 1245 CTGATGACACACCCCATACATGAGCGCTGGGACTTTTGAGCAGCATTATTGTGCC 1304
 Db 591 CysValasp-----HisSerProGluPro-----Ser 599
 QY 1305 CCAGCCAGTTCAGGACCTCCAAACTGAGAAATAGCAGGCGCTCTCTGTATC 1364
 Db 600 LeuAsnCysThrIleGlyAspProAlaLeuAsnGlyLysLeuGluGluPheSerCysVal 619
 QY 1365 TCCTGGGAAGTAGTGAGGCCACATGTGCTGCTGCTGGCCAGGCTGAGGAACGACAG 1424
 Db 620 GlyGlyArgLeuCysProLeuAsnThrAlaCysLeuAlaAlaProSer--GlySerProA 639
 QY 1425 CAGGAGCACTGTCTCCCAAGCTTGCTGAGCAGATGATCCTGGAGGAGTTTGGAAAGTGC 1484
 Db 639 laValCysCysTyr-----ArgProP 646
 QY 1485 CTGTCGCGAGATCTCCACATCGATTCAAGTCCAGGGGCTGAAATTTGAGTAGGCTGCA 1544
 Db 646 roValAlaIleThrProAlaProThrThrVal-----656
 QY 1545 AGCTGCTAAAGGGAATGCTGTGCAAGCCTCAGCCTGGGAATCTGCACCGAGGAAGC 1604
 Db 657 -----ProIleProValSerThrAla--AlaProTh 666
 QY 1605 TGGTCCCGCAGGAGGAGGCGCGCATCATGCT-----CAGGT 1646
 Db 666 rSerAlaProGlyProValAspProCysGlnProGlyValAsnValAsnCysThrLysG 686
 QY 1647 CAGCTGTGAGTCTGAGTGAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1706
 Db 686 yGluCysArgLeuValAlaIleCysLys--TyrTrpProCysPheAla-----701
 QY 1707 TCGAAGCAGGAGGAGCCCTGGGGATGGACACTGGCCCTCTCTCTCTCTCTCTCTCTCTCT 1766
 Db 702 -----LeuProThrCysValAspProSerProProSerProProValGluCysP 717
 QY 1767 TCTGGGAGTCTCTCCCTGTCATAAAGAGGAGGCTGATTTCTACCTGTGTGTATGT----- 1822
 Db 717 roValGlyLysProAlaLeuAspGluLysLeuGluGluPheSerCysLysAspCysProp 737
 QY 1823 -----GTTTGTCTTTCAATTTGCTTAGTACTACTCTCATTCAGTTATTATGAGCCAG 1874
 Db 737 heAsnThrValCysTyrLys-----743
 QY 1875 CCTCAAGTTAGAGCTAGGCTCTCTCTCAGTGGAGT-----CTGCCCAATACATA 1928
 Db 744 -----GlyAlaValCysCysValProTyrSerGlyAsnArgProSerGlyProA 760

QY 1929 CAAGTCAGTGGCCCATCAGGGGTTTTTCCAGGCCAGGCTGTGACAGGAGATATGGAGG 1988
 Db 760 laGlyProAlaGlyProAlaGlyProGluArgPro-----771
 QY 1989 GGGTCGGGTTAGAGCTGGGTTTGTTCGATTTTTCGATTTTTCCTCTCTCTCTCTCT 2048
 Db 772 -----AlaThrSerVal-----775
 QY 2049 GCTTGAAGTGAGAAACTGTCTCTGCTCCAACTTTCATCAATTAATCTCTCTCTCTCT 2108
 Db 776 -----ProLeuAspProCysThrProGlyLeuAsnValAsnCysThrSerG 791
 QY 2109 TCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2168
 Db 791 lyValCys-----ArgLeuValGluAspCysArgArgProGly-----CysP 805
 QY 2169 CACACTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2200
 Db 805 roAlaValProThrCysIleAspArgAspPro 815
 RESULT 7
 G01763
 acrophin-1 - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
 C:Accession: G01763
 R:Margolis, R.L.
 submitted to the EMBL Data Library, March 1995
 A:Reference number: G08343
 A:Accession: G01763
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1184 <MAR>
 C:Cross-references: EMBL:U23851; NID:g915325; PID:g915326
 C:Genetics:
 A:Gene: GDB:DRPLA; B37
 A:Cross-references: GDB:270336; OMIM:125370
 A:Map position: 12p-12p
 Alignment Scores:
 Pred. No.: 0, 0528 Length: 1184
 Score: 125.50 Matches: 126
 Percent Similarity: 29.91% Conservative: 43
 Best Local Similarity: 22.30% Mismatches: 206
 Query Match: 3.10% Indels: 191
 DB: 2 Gaps: 24
 US-09-743-237-1 (1-2241) x G01763 (1-1184)
 QY 1991 CCCCCTCCCATATCTCTGTC-----ACAGGCGCTGGCTGGAACACCCCTGA 1944
 Db 253 ProProProThrThrProIleSerValSerSerSerGlyAlaSerGlyAlaProProThr 272
 QY 1943 TGGCCACCTGACTGTATGTATTTGGCAGATCCACTGGAAGAGCCCTAGCTCTCT 1884
 Db 273 LysProProThrThrProVal--GlyGlyLysLeuProSerAlaProProProAla-- 291
 QY 1883 AACTTGGAGCTGGCTCATAATAACTGAATGAGGAGTACTACTAAGCAATTTGAAAGCAA 1824
 Db 291 -----291
 QY 1823 CACATAACAACAGTAGAAATACCCCTCTTTTATGAGGCGAGGAGTCCCCAGAGG 1764
 Db 292 -----AsnPheProHisValThrProAsnLeuProProProAl 305
 QY 1763 CCCCCCAGGACAGGAGGCGCAGTGTCCATCCCGGAGGTGCTCTGCTGCTTCCACTT 1704
 Db 305 aLeuArgProLeuAsnAsnAlaSerAlaSerProProGlyLeuGlyAla-----321
 QY 1703 TAGGATACCTTTCAGTAGGCTGGCCAGTAGTACCATGACAGTCACTCAGACCTCACAGC----- 1649
 Db 322 -----GlnProLeuProGlyHisLeuProSerProHisAlaMetGl 335

[illegible]

```

Db      476 aGlyAsnProSerLeuHisGly-::: ||| :::: |||
QY      1588 TTCCAGGCGCTGAGGCTTCCACAGCAGCATTCCTCCCTTACACAGCTGCGCTACTCAATT 1529
Db      491 nSerGln-----ProHisVal-----ArgAsnGlyAspAsnII 502
QY      1528 TTACGCCCCCTGGACTTGAACTCGATGTGGAGAATCTCGACAGCAGCTTTC----- 1477
Db      502 eSerAlaPro-----AsnValLysAlaSerGluSerGlyHisPheAlaSerTh 518
QY      1476 -----CAAACCTCCAGGATCA---TCTGCTCAGCCAAAGCTTGGGAA 1436
Db      518 rArgGlnAsnLysProGlnProProAlaSerIleSerValValPro----- 534
QY      1435 CAGTGCCTCTCTCTCTCTCCCTCCCTGCGCAGCAGCAGCATGTGCGCTCCACT 1376
Db      535 ----AlaProAlaPheIle-----TyrProAlaAsnHisLeuGlnProValMe 550
QY      1375 ACTTCCAGGAGATACAGAGAGAGCGCTCCCTATTCTTCTCAGTTTGGAGCTCCTGAG 1316
Db      550 tValProSerLys---SerSerArgPro----- 558
QY      1315 AACTTGGCTGGGACAAATAATGCTGCTCTCAAAGTCCCAAGGCTCCATGTAGTGGGT 1256
Db      559 -----ThrLysSerProHisLeu---AlaValGlyLe 568
QY      1255 GTGCTCATCAGCATTTTTCGTTCTGCGACTTCTCTCATAGTTTTCGAAGCAATGCTATTG 1196
Db      568 uAlaSerAlaAsnPhe----- 573
QY      1195 CAAATGGAGAACACATGATTTTGGCCTCATAGCCTACACAGTAGTCTTCTCAGGAGC-- 1138
Db      574 -----SerHisProSerSerAlaSerGln 582
QY      1137 -----CTGAGCGCTTACAGTTTCACCTTTCGCTG 1109
Db      582 uValSerSerProTyrLeuThrValIleProAsnAsnAlaTyrSerPheGlnLeuSerSe 602
QY      1108 TGTGCAAGTTTACGAGCTCCACAGCGCTTTCCCTTTTGGTTGGTGAAGCTT----- 1054
Db      602 rThrIleArgGlyGlyThrProSerGlnAlaValProPheTyrAsnGlySerPheTyrSe 622
QY      1053 -CAGGATTTCTATCAAGACACCCCTTTATGGCTT-----TGAAGCGCTCGAGC 1007
Db      622 rProGlnMetPheGlnGlnProProGlnProLeuGlnArgGlnSerGlnAlaGlnArgGln 642
QY      1006 TCATGGCGCAGGTTTTCAGCTGCGAGCTGTTGCGAGAAGTCCCGCTGGAGNAGCAGTCA 947
Db      642 uSerLysAlaSerSerCysSerSerSerHisArgGlnPro----- 656
QY      946 CAGTACCCAGACAGAGTATTATTGTTGGTGGCCCTTGCAGAGCAGGTCCAGAGGAAAGGCA 887
Db      657 -----GlnValSerValAsnSerLeuSerGlnAlaAsnValGlnHisArgGln 674
QY      886 GCCCATTTGACAGCTGATGGGAGAGCTCCATATCAACCTGTGTGATGAGATTAACTTC 827
Db      674 n----- 674
QY      826 GCTGTACTGGCAGAAGTCTTCCCACTAAAGTTGTTCATTTAGGCTTACGCTCTGACTCA 767
Db      675 -----MetSerGlnLysPheGluValAla-----GlyValAsnThrAs 687
QY      766 GGGAAATGCTCTCTGCTCATCGTACTGAGGAAGCAGATGGAGCGCTTTGAGCTCCCTCGCG 707
Db      687 pSerArg-----GlySerHisThrGlnLysGlyGly----- 697
QY      706 CCACAGTTGCTATGACAGAGCATCTGGGCGCTCTTTCAGCTGACAAATCACCATTGGGG 647
Db      698 -----ProPheGlyGlnIleMetAlaAlaProValGlnProGlnAs 711
QY      646 CTGAGCTCTTCTCT-----TCCGAGGCGAGCTGGAGGCTCCT----- 610
:::|||||::: ||| ||| |||||

```

```

Db      711 nPheSerMetSerPheAlaSerIleAlaSerSerAlaProProAlaThrLeuAsnPheSe 731
QY      609 -----CTGCCTCTCGGATGATG----- 592
Db      731 rSerAsnGlyTyrHisIleSerThrProGlyValAlaHisGlnLysAsnHisGlnSe 751
QY      591 -----GGAACCTTGCAACAGGATTCTCTGAGCCAGAGGCGCCCTGG 554
Db      751 rSerGluAlaLysThrGlyGlySerCysSer-----SerAsnAlaGluAspPro-- 768
QY      553 AAAGCTGCTCTTTCAGGCG-----TGGCGCTGCGACACTACCACTGCTCTTTTGATT 500
Db      769 -LysLysAsnLeuGlnGlyLysProGlnGlyMetMetAsnGlyHisThrLeuVal-PheA 788
QY      499 TCAACACATTTTCTTCTGCTTGTGCTTGGGAACCTGGCAGTAGTTTCATCGGAGTCT 440
Db      788 spAsnProSerArgThrLeuAsnPheValSerGlyThrTrpPro----- 802
QY      439 CACCAAAATGATTTCCTCCCTGTA----- 417
Db      803 -----ProProAlaAlaThrAlaIleAsnGlyAspProSerValPheThrG 818
QY      416 --CACCTAGCACTCAGGATGTAGAGCAGGAGAAATCAGAGGTTCCTCAAGTTATATAGC 359
Db      818 InHisLeuThrGlnArg-----GlnGlnGlnSerGlyArgSerLysMetMetThrH 836
QY      358 AATTTCAGGAATTGCTATATAGCTTATGCAATATGTTGAGAAGCAAGAAATCATCAAAAC 299
Db      836 isSerGlnAlaAspSerValSerAlaThrSerSerGlnTrpLysAsnProAlaThrSer- 855
QY      298 TCTACATAGGTTTAAAGAACTGTCCTTATCTTACAGAGGGGGAGATCATCAACTGAAA 239
Db      856 ----SerSerLeuThrSerCysThrSerLeuAsnLeuLysGlnPheGlnSerGlnGlnG 874
QY      238 AAATCCTCAATAATGGAAGTCTTTTGGTGTGAAGAAAAAATAATCCAGCAACCCCAA 179
Db      874 InIleArgThrHisGlyGlnThrGlnIleSerPheAlaAlaProThrAsnProGlnPro 894
QY      178 ACTACAGGCTGAAGACTATCTGCAAGAACACAGCAAA---AAGGCCCTGGGTCTATCCCTA 122
Db      894 er-----GlnGlyLysGlnGlyArgSerGlyGlySerSerPros 907
QY      121 GC-----CCTGTGGGGCGGAGGTGGTGGAGTCACAA 89
Db      907 erValThrGlySerAlaSerHisGlyLysProAlaAsnSerLysValSerAsnSerLysA 927
QY      88 GT-----CCACAGCGCCAGGAACACACA 66
Db      927 IaLeuLeuSerProValProLeuSerGlnGluHisThr 940

```

```

RESULT 11
S50832
atrophin-1 - human
C.Species: Homo sapiens (man)
C.Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
C.Accession: S50832
R.Nagafuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.;
Nature Genet. 8, 177-181, 1994
A.Title: Structure and expression of the gene responsible for the triplet repeat diso
A.Reference number: S50832; MUID:95144175; PMID:7842016
A.Accession: S50832
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-1184 <NAB>
A.Cross-references: EMBL:D31840
C.Genetics:
A.Gene: GDB:DRPLA; B37
A.Cross-references: GDB:270336; OMIM:125370
A.Map position: 12p13.31-12p13.3112p-12p
Alignment Scores:
Pred. No.: 0, 165 Length: 1184
Score: 119.50 Matches: 125

```


C;Function:

C;Function: control of embryonic development by tissue- and stage-specific regulation
A;Description: homeotic protein Hox D4; homeobox homology
C;Superfamily: homeotic protein Hox D4; homeobox homology
C;Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation
F;210-366/Domain: homeobox homology <Hox>

Alignment Scores:		
Pred. No.:	0.218	Length:
Scores:	17.50	Matches:
Percent Similarity:	36.36%	Conservative:
Best Local Similarity:	29.29%	Mismatches:
Query Match:	2.90%	Indels:
DB:	1	Gaps:
		309

US-09-743-237-1 (1-2241) X S09257 (1-309)

1829	Qy	AGCAACACATACACAAAGTAGAAATCACCCTCTTTTATCAGCGCAGGAGTCCCC	1770
24	Db	ThrGlnHisSerGlySerAlaGlySerSerAlaSerTyHisProHisProHisPro 43	
1769	Qy	AGAGGGCCTCCCGCAGGACAGGAGGCCAGTGTCATCCCGAGGTGGCTCTGCCTGCTT	1710
44	Db	HisAlaProProPro-----Pro-ProProProProProHisLe 57	
1709	Qy	GCATTTTAGGATACCTTAGCTAGGCTGGCCAGTACCATGCAGATCACTCAGACGCTCACAG	1650
57	Db	uHisAlaAlaHisPro-----GlyProAlaLeuProGluTyPheProArgProArg 75	
1649	Qy	CTGACCTGGCCATGATGCGGGCCTCTGCTCTCCCTGGGCACACAGCTTCTCGGTGCAG	1590
75	Db	gGluProGlyTyGlnAla---ProAlaAlaProProGlyProProGlyProPro----- 92	
1589	Qy	ATTTCACAGGCTCAGGCTGCCACAGCATTCCTCCCTTACCAGCTTGCACGCTACTCAAT	1530
93	Db	-----ProGluAlaLeuTyProAlaGlnAlaProSerTyProGlnAlaProTySer-T 111	
1529	Qy	TTTCAGCCCTTGACTTGAACTCGATGTGGAAATCTGCGACAGCAGCCTTCCAAATC	1470
111	Db	yrSerSer-----AlaGlySerAlaAlaProGlyProGluGlnProp 125	
1469	Qy	CTCCAGGATCATCTGCTCAGCCAAAGCTGGGGAAACAGTGCTCTGCTCTGCTTCTCACC	1410
125	Db	roProGlyAlaSerProProPro-----ProProProAlaLysGlyHisP 140	
1409	Qy	CTGGGCCACAGCGGCACATCTGGCCTCCACTACTTCCC-----AGGA 1365	
140	Db	roGlyProAla-----GlnProLeuLeuProGlyHisAlaLeuGlnArgA 155	
1364	Qy	GATACAGGAGAGGCCTGCCTATTTTCTCAGTTTGGAGGCTTGAGAACTTGGCTGG	1305
155	Db	rgCysGluAlaAlaProAla-----AlaGlyAlaG 165	
1304	Qy	GGACAATAATAGCTGCTCTCAAAGTCCCGAGCTCCATGTAGTGGGGT 1255	
165	Db	lvThrGlyProGlyCysAlaLeuLeuProAspLysSerLeuProGlyLeu181	

RESULT 13

MEGF6 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: Tl3954
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: Tl3954
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1574 <NA>
A:Cross-references: EMBL:AB011532; NID:g3449293; PIDN:BAA32462.1; PID:g3449294
A:Experimental source: strain Sprague-Dawley; brain
C:Genetics:

A:Accession: A59110
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-2767 <MAL>
A:Cross-references: GB:X05615; NID:g37173; PIDN:CAA29104.1; PID:g37174
A:Note: revision to S00014
R:Walthery, Y.; Lissitzky, S.
Eur. J. Biochem. 165, 491-498, 1987
A:Title: Primary structure of human thyroglobulin deduced from the sequence of its 84
A:Reference number: S00014; MUID:87246630; PMID:3595599
A:Accession: S00014
A:Molecule type: mRNA
A:Residues: 1-1041, 'Y', 1043-1057, 'T', 1059-2767 <MAL2>
A:Cross-references: GB:X05615; NID:g37173
A:Note: this sequence is revised in A59110
R:Walthery, Y.; Lissitzky, S.
Eur. J. Biochem. 147, 53-58, 1985
A:Title: Sequence of the 5'-end quarter of the human-thyroglobulin messenger ribonuc
A:Reference number: A01532; MUID:85127024; PMID:3971976
A:Accession: A01532
A:Molecule type: mRNA
A:Residues: 1-730 <MAW>
A:Cross-references: GB:X02154; NID:g37175; PIDN:CAA26089.1; PID:g1335349
A:Note: the translated sequence in GenBank entry HSTHYRR5, release 111.0, (PIDN:CAA260
n 1-Met
R:Parma, J.; Christophe, D.; Pohl, V.; Vassart, G.
J. Mol. Biol. 196, 769-779, 1987
A:Title: Structural organization of the 5' region of the thyroglobulin gene. Evidence
A:Reference number: S03422; MUID:88062712; PMID:3681978
A:Accession: S03422
A:Molecule type: DNA
A:Residues: 1-134, 'G', 136-415; 640-652, 'G', 654-733, 'A', 735-737; 880-983, 'DR', 985-999 <P
A:Cross-references: EMBL:X06059; NID:g37145; PIDN:CAA29454.1; PID:g1359884; EMBL:X060
R:Christophe, D.; Cabrer, B.; Bacolla, A.; Targovnik, H.; Pohl, V.; Vassart, G.
Nucleic Acids Res. 13, 5127-5144, 1985
A:Title: An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream
A:Reference number: I38343; MUID:85269632; PMID:2991855
A:Accession: I38343
A:Status: translation not shown; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-22, 'GRF' <CHR>
A:Cross-references: EMBL:X02749; NID:g37162; PIDN:CAA26527.1; PID:g758106
R:Targovnik, H.M.; Cochaux, P.; Corach, D.; Vassart, G.
Mol. Cell. Endocrinol. 84, R23-R26, 1992
A:Title: Identification of a minor Tg mRNA transcript in RNA from normal and goitrous
A:Reference number: I57669; MUID:92347597; PMID:1639210
A:Accession: I57669
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1503-1508, 'L', 1567-1601 <REN>
A:Cross-references: GB:S40807; NID:g252170; PIDN:AB22685.1; PID:g252171
A:Note: this sequence fragment represents a minor splice form
R:Marriq, C.; Lejeune, P.J.; Venot, N.; Vinet, L.
FEBS Lett. 242, 414-418, 1989
A:Title: Hormone synthesis in human thyroglobulin: possible cleavage of the polypepti
A:Reference number: S02266; MUID:89121111; PMID:2914619
A:Contents: evidence for Tyr-149 as ring donor for thyroxine formation
A:Accession: S02266
A:Molecule type: protein
A:Residues: 101-109, 114-121; 126-131; 143-149 <MAR>
A:Note: only the first peptide was sequenced; others were isolated and their amino ac
R:Gentile, F.; Salvatore, G.
Eur. J. Biochem. 218, 603-621, 1993
A:Title: Preferential sites of proteolytic cleavage of bovine, human and rat thyroglob
A:Reference number: S39431; MUID:94094855; PMID:8269951
A:Accession: S39431
A:Molecule type: protein
A:Residues: 20-27; 532-541; 547; 616-619, 'X', 621; 1001-1005; 1009-1011, 'X', 1013; 1424-1
R:Xiao, S.; Pollock, H.G.; Taurog, A.; Pawlitch, A.B.
Arch. Biochem. Biophys. 320, 96-105, 1995
A:Title: Characterization of homonogenic sites in an N-terminal, cyanogen bromide fr
A:Reference number: S66241; MUID:95314327; PMID:7793989
A:Accession: S66241

Alignment Scores:	
Pred. No.:	0.309
Score:	116.50
Percent Similarity:	32.42%
Best Local Similarity:	20.84%
Query Match:	2.91%
DB:	1
Length:	2767
Matches:	144
Conservative:	80
Mismatches:	269
Indels:	199
Gaps:	35

Qy		447	GATGAAC	TACTGCCAGGTTCCCAAGGGCAGCAAGCAAGCAAAAAAGTGTGAAATCAA	506
			:::::	:::::	
Db		1132	GlUGLUUUA	rqPProGLIVS	SerSerAlaGLnCysProSerLeuCvAsnValLeu
			:::::	:::::	:::::

QY 507 GAAGCAGGT-----GGTAGTGTGCCAGGGCGAGCCCTGAA 542

Db	1152	LySerGlyValLeuSerArgArgValSerProGlyTyrValProAlaCysArgAlaGlu	1171
QY	543	GAGCAGCTTTC-----CAGGCCCCCTCGCTCAGGAATCTGT---TCGAAGTTC	590
Db	1172	AspGlyGlyPheSerProValGlnCysAspGlnAlaGlnGlySerCysTrpCysValMet	1191
QY	591	CCATCATCCACGAG-----CGACGAGGAGGCC	617
Db	1192	AspSerGlyGluGluValProGlyThrArgValThrGlyGlyGlnProAlaCysGluSer	1211
QY	618	TCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTGATTTGTCAGCTGAAGAGGCC	677
Db	1212	ProArgCysProLeuProPheAsnAlaSerGluValVal-----Gly	1225
QY	678	GCCACAGATCTGC-----ATAGACAACCTGTGGCGCGAGGAGCTCAAAGCGCTCCAT	731
Db	1226	GlyThrIleLeuCysGluThrIleSerGlyProThrGlySerAlaMetGlnGlnCysGln	1245
QY	732	CTGCTTCTCAGTACGATGACACAGACAGTTTCCCT-----	767
Db	1246	LeuLeuCysArgGlnGlySerTrpSerValPheProGlyProLeuIleCysSerLeu	1265
QY	768	-----CAGTCAGAGTCCCTTAAGCAATGACACACTTGTGAGGAGCTTCCCA	869
Db	1266	GluSerGlyArgTrpGluSerGlnLeuProGlnProArgAlaCysGlnArgProGlnLeu	1285
QY	813	---CTGCCAGTACCAGGAAGTTAAATCTCATCACACAGGTTGATATGGAGCTTCCCA	869
Db	1286	TrpGlnThrIleGlnThrGlnGlyHisPheGlnLeuGlnLeuProGlyGlyMetCys	1305
QY	870	TCAGCTGTCAATGGGCTGCCTTCCCTCTGGACCTGCTCTGCAAGGGCCACCAAAVA	929
Db	1306	SerAlaAspTyrAlaGlyLeuLeuGlnThrPheGlnValPheIle---LeuAspGluLeu	1324
QY	930	ACTCTGCTGGTACTGT-----GACTGCTTCTCCAGCGGGGACTTCTGCAACAGCTGC	983
Db	1325	ThrAlaArgglyPheCysGlnIleGlnValIlyThrPheGlyThrLeuValSerIlePro	1344
QY	984	AGTGCACAACCTGGCC-----	1001
Db	1345	ValCysAsnAsnSerSerValGlnValGlyCysLeuThrArgGluArgLeuGlyValAsn	1364
QY	1002	-----CATGAG	1007
Db	1365	ValThrTrpLysSerArgLeuGluAspIleProValAlaSerLeuProAspLeuHisAsp	1384
QY	1008	CTCAGCGCTTCAAGCCATAAAGCGGTGCTTTGATAGAAATCCTGAA-----	1055
Db	1385	IleGluArgAlaLeuValGlyLysAspLeuLeuGlyArgPheThrAspLeuIleGlnSer	1404
QY	1056	---GCTTTCACACAAAATGGGAAAGGCGTCTGGAGCTCTAAACTTCACACACAGC	1112
Db	1405	GlySerPheGlnLeuHisLeuAspSerLysThrPheProAlaGluThrIleArgPheLeu	1424
QY	1113	AAAGGTGCAACTGT-----AAGCGCTCAGGCTCGCTGAAGAACTAC---	1154
Db	1425	GlnGlyAspHisPheGlyThrSerProArgThrArgPheGlyCysSerGluGlyPheTyr	1444
QY	1155	-----TGTGAGTGCTATGAGGCC	1172
Db	1445	GlnValLeuThrSerGluAlaSerGlnAspGlyLeuGlyCysValLysCysHisGluGly	1464
QY	1173	AAATCATGTGTTCTTCCATTTCCAAATGATGCTTGTGCAAA-----ACTATGAAGAA	1226
Db	1465	SerTyr-----SerGlnAspGluGluCysIleProCysProValGlyPheTyrGlnGlu	1482
QY	1227	AGTCCAGAACGAAAATGCTGATGACACACCC-----CACTACATGGAGCGCTGGG	1277
Db	1483	GlnAlaGlySerLeuAlaCysValProCysProValGlyArgThrThrIleSerAlaGly	1502
QY	1278	GACTTTGAGACGACCATTTATTTGTCCTCCA-----	1307

Db 1503 AlapheSerGlnThrHisCysValThrAspCysGlnArgAsnGluAlaGlyLeuGlnCys 1522
QY 1308 -----GCCAAGTTCTCAGGACCTCCAAAAGTGCAGAAAAATAGCAGGCTTCTCC 1358
Db 1523 AspGlnAsnGlyGlnTyArgAlaSerGlnLysAspArgGlySerGlyLysAlaPhe--- 1541
QY 1359 TGTATCTCC-----TGGGAAGTAGTGGAGGCC-----ACA 1388
Db 1542 CysValAspGlyGluGlyArgArgLeuProTrpTrpGluThrGluAlaProLeuGluAsp 1561
QY 1389 TGTGCTGCTGCTGGCCAGCGGTGAGAA----- 1418
Db 1562 SerGlnCysLeuMetGlnLysPheGlnLysValProGluSerLysValIlePheAsp 1581
QY 1419 -----GCAGACGAGCAGCAGCTGTTCCTCCCAAGCTTGGCTGAGCAGATGATC 1463
Db 1582 AlaAsnAlaProValAla-ValArgSerLysValProAspSerGluPhePro-----Va 1599
QY 1464 CTGGAGGAGTTTG-----GAAGGTGCTGTGCGCAGATTCTCCACATCGAG 1508
Db 1599 lMetGlnCysLeuThrAspCysThrGluAspGluAlaCysSerPhePheThrValSerTh 1619
QY 1509 TTCAAGTCCAAAGGGCTGAAATTGAGTAGCTGCAAGCTGTTAAAGGGGAATGC----- 1563
Db 1619 rThrGluPro-----GluLeSerCysAspPh 1628
QY 1564 -----CTGTGCAAGCTCAGCCCTGGGAATCTGCACC-----GAGGAAGCTGTGCGCCA 1613
Db 1628 eTyAlaTrpThrSerAspAsnValAlaCysMetThrSerAspGlnLysArgAspAlaLe 1648
QY 1614 GGGAGGACGAGCGCGGCATCATGCGCAGCTGAGCTGAGCTGAGTGAT-CTGCA 1672
Db 1648 uGlyAsnSerLysAlaThrSerPheGlySerLeuArgCysGlnValLysValArgSerHi 1668
QY 1673 TGGTACTGGCCAGCTACT-----CAAGGTATCTCTAAAGTCAAGCA 1714
Db 1668 sGlyGlnAspSerProAlaValTyLeuLysGlyGlnGlySerThrThrLeuGl 1688
QY 1715 GGCAGAG---CCACCCTGGGATGACACACTGGCCCTCCTGCTGGGAGGCC----- 1766
Db 1688 nLysArgPheGluProThrGlyPheGlnAsnMetLeuSerGlyLeuTyArgProIleVa 1708
QY 1767 -----TCTGGGACTCCCTGCC-----TCGATAAAAGAGGGTG 1801
Db 1708 lPheSerAlaSerGlyAlaAsnLeuThrAspAlaHisLeuPheCysLeuAlaCysAs 1728
QY 1802 ATTTTCTACTTGTGTATGTTGTTGCTTCAAAATTGCTTAGTAGTACCTCCATTCAAGT 1861
Db 1728 pArgAspLeuCysCysAspGlyPheValLeuThrGlnValGlnGlyAlaIleIleCy 1748
QY 1862 T---ATTATGACCCAGCTCAAGTTAGAGAGTAGGCTCTTCTTCAGGTGGACTCTGCC 1918
Db 1748 sGlyLeuLeuSerSerProSerValLeu---LeuCysAsnValLysAspTrpMetAspPr 1767
QY 1919 AAATCACAATACAGTACAGTGGCCATCAGGGGTTTTTCCAGGCCAGGCTGTGACAGAG 1978
Db 1767 oSerGlu-----AlaTrpAlaAsnAlaThrCysProGlyValThrTyAspGlnGl 1784
QY 1979 ATATGGAGGGGGTTCGGTTAGAGCTGGGT 2009
Db 1784 u---SerHisGlnValIleLeuArgLeuGly 1793

Search completed: July 24, 2003, 14:35:40
Job time : 141.48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:58:44 ; Search time 81.9566 Seconds
(without alignments)
6494.675 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 tatcctgtgtggtgcccgc.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Fgapop 10.0 , Fgapext 0.5

Delop 6.0 , Delext 7.0

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 903798

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet_n2p_model -DEV=xlp

-Q=/cgn2_1/USFTO.spool/US09743237/runat_24072003_113106_17768/app_query.fasta_1.4750

-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosu62

-TRANS-human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100

-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US09743237@cgn.1.1.282@runat_24072003_113106_17768

-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100

-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
------------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

US-09-764-891-3988

: Sequence 3988, Application US/09764891

: Publication No. US20030077808A1

: GENERAL INFORMATION:

: APPLICANT: Rosen et al.

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

: FILE REFERENCE: PC006

: CURRENT APPLICATION NUMBER: US/09764.891

: CURRENT FILING DATE: 2001-01-17

: Prior application data removed - consult PALM or file wrapper

: NUMBER OF SEQ ID NOS: 10231

: SOFTWARE: PatentIn Ver. 2.0

: SEQ ID NO 3988

: LENGTH: 147

: TYPE: PRT

: ORGANISM: Homo sapiens

: FEATURE:

: NAME/KEY: SITE

: LOCATION: (146)

: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-891-3988

Alignment Scores:

Sequence 3988, Ap
Sequence 13, Appl
Sequence 6932, Ap
Sequence 18, Appl
Sequence 20, Appl
Sequence 8, Appl
Sequence 14, Appl
Sequence 4, Appl
Sequence 5, Appl
Sequence 6, Appl
Sequence 16, Appl
Sequence 24, Appl
Sequence 10, Appl
Sequence 12, Appl
Sequence 2, Appl
Sequence 142, App
Sequence 1116, Ap
Sequence 1116, Ap
Sequence 50, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 96, Appl
Sequence 2, Appl
Sequence 240, App
Sequence 16, Appl
Sequence 6, Appl
Sequence 8, Appl
Sequence 12, Appl
Sequence 5, Appl
Sequence 20, Appl
Sequence 20, Appl
Sequence 282, App
Sequence 22, Appl
Sequence 36, Appl
Sequence 357, App
Sequence 14, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 1858, Ap
Sequence 10, Appl
Sequence 1857, Ap
Sequence 1856, Ap
Sequence 38, Appl
Sequence 36, Appl
Sequence 34, Appl

Alignment Scores:		
Pred. No.:	0.0178	Length: 1192
Score:	131.50	Matches: 135
Percent Similarity:	28.72%	Conservative: 56
Best Local Similarity:	20.30%	Mismatches: 215
Query Match:	3.28%	Indels: 260
DB:	15	Gaps: 36

Alignment Scores: 0.0178 Length: 131.50 Matches: 28.77% Conservative: 20.30% Mismatches: 20.30% Indels: 1.28% Gaps: 3.25%

Pred. No.: 0.0178 Length: 131.50 Matches: 28.77% Conservative: 20.30% Mismatches: 20.30% Indels: 1.28% Gaps: 3.25%

US-09-743-237-1 (1-2241) x US-10-189-971-18 (1-1192)

717	CTCAAAGCGCTCATCTCTCTCTCCTCAGTACGATGACACAGACGAGTTCCTCCTCAGTCAGAG	776
QY		
717	CTCAAAGCGCTCATCTCTCTCTCCTCAGTACGATGACACAGACGAGTTCCTCCTCAGTCAGAG	776
Db		
381	-----TyrLeuGlycylSerTyrLeuSerAsnGlnIuPheProAspProArg	396
QY		
381	-----TyrLeuGlycylSerTyrLeuSerAsnGlnIuPheProAspProArg	396
Db		
777	CTCCCTAAGCCCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCACGCGAAGTTAAAT	836
QY		
777	CTCCCTAAGCCCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCACGCGAAGTTAAAT	836
Db		
397	GluProCysAsnLeuCysThrCysLeuGluGlyIuPhe-----	408
QY		
397	GluProCysAsnLeuCysThrCysLeuGluGlyIuPhe-----	408
Db		

837 CTCATCACACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCAATGGGGCTGCCTTTCCC 896

Db	-----ValThrCysGlyArgProCysGluProProGlyCysSerHisPro	409
QY	TCTGGACCTGCTCTCAAGGCCACCACCAATAACTCTCTCTGGGTACTGT	897
Db	LeuIlePro-----SetGlyHisCysCysProThr	425
QY	-----GACTGC-----TTTCCAGCGGGAC	948
Db	CysGlnGlyCysArgTyrrHisGlyValThrAlaSerGlyGluThrLeuProAspPro	435
QY	-----TTTCTGCAACAGCTGCAGCTGCACAACACCTCGCGCATCAGCTCGAGCGCTTC	969

Db	455	LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluuHisGln-----	470
QY	1020	AAAGCCATAAAGCGGTGCTTGATAGAAATCCTGAAGCTTTCCCAACCAAAAATGGGAAA	
Db	471	-----AspGlyGluGluPheGluGluProAlaGlySer	481
QY	1080	-----CGCCGCTCTGGGAGCTGCTAAACTTCGA-----	110
Db	482	CysGluTrpCysArgCysGlnAlaGlyValSerCysValArgLeuGlnCysProPro	501
QY	1107	-----CACAGCAAGGGTGCACTGTAAAGCGC---TCAGGCTGC	114
Db	502	LeuProCysLysLeuGlnValThrGluArgGlySerCysProArgCysArgGlyCys	521

QY 1143 CTGAAG----- 1148
 Db 522 LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys 541
 QY 1149 AACTACTGTGAGTGTATCAG-----GCCAAATCATGTGTTCTCCATT 1193
 Db 542 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 561
 QY 1194 TCAGAA-----TGCATTGCTTGCAGAAATATATGAGAA 1226
 Db 562 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 581
 QY 1227 ACTCAGAACGAAAAATGCTGTATGAGCACACCCACTACATGAGCGCTGGGAC---TTT 1283
 Db 582 GluGlyArgLys-----TyrGluProGluGluSerPhe 592
 QY 1284 GAGAGCAGCCATTATTGTCGCCAGCCAGTTC-----TCAGGA 1322
 Db 593 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 610
 QY 1323 CCTCCAAACTGAGAAAAATAGCGAGCGCTTCTCTGTATCTCTGGAAGTAGTGAG 1382
 Db 611 ProProSerLeuArgCysHisArg-----GlnCysProSer-----LeuValGly 626
 QY 1383 GCCACATGTGCTGCTGCTGCCGCCAGGTGAGGAGCAGCAGCAGCAGTGTCCCA 1442
 Db 627 CysProProSerGlnLeuLeuProGly-----ProGlnHisCysCysPro 642
 QY 1443 ACCTTGGCTGACCATGATCC----- 1464
 Db 643 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 662
 QY 1465 -----TGCAGGAGTTTGGAAAGTGC 1484
 Db 662 oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 682
 QY 1485 CTGTGCGAGATTCTCCATCATCAGTTCA-----AGTCCAGGGGCTGAAATAG 1535
 Db 682 acysProGluLeuSerCysProLeuSerGluArgHisTrpProGly----- 698
 QY 1536 TAGCGTGAAGTGTGTAAGGGGAATGCTGTGGAAGCCTCAGCCCTGGGAATCGCAC 1595
 Db 699 -SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs 718
 QY 1596 CGAGGAAGTGTGTCGCCAGGAGGA-----GCAGAGCGCGGCATCATGCGCA 1643
 Db 718 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHi 738
 QY 1644 GGTACGCTGTGA-----GGTCTGAGTATGTCATGTGCTAGTGGCCAGCC 1687
 Db 738 sValGluCysHisLeuGluGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 758
 QY 1688 TACTCAAGGTATCCTAAAGTGAACGACGAGCAGCCAGCCCTGG----- 1730
 Db 758 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisGl 778
 QY 1731 -----GGATGGACACTG----- 1742
 Db 778 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 798
 QY 1743 -----GCCCTCCCTGCTCCCTGGGAGGCC 1765
 Db 798 sMetalGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 816
 QY 1766 CTCTGGGACTCCCTGCTGATCAAAAAAGAGGGTGATTTCTACTTGTGTATGTGT 1825
 Db 817 -----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 830
 QY 1826 TCTTTCAAAATTGCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCAGCCCTCAAGTTAG 1885
 Db 831 -----ArgCysLeuProArgProAlaSerCysMetalAlaPheGlyAspProHisTyrAr 848
 QY 1886 AGAGCTAGGCTCT-----TCTTCAGGTGACTCTGCCCAATACATACAGTACAG 1936

Db 848 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 868
 QY 1937 GTGGCGATTGAGGTTTTCAGGCGAGGCTGTGACAGGATATGGAGGGGGTTCGG 1996
 Db 868 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 887
 QY 1997 GTTAGAGCTGG 2007
 Db 887 LyValAlaTrp 890
 RESULT 5
 US-10-189-971-20
 ; Sequence 20, Application US/10189971
 ; Publication No. US20030028907A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Scoville, John
 ; APPLICANT: Turner, C. Alexander Jr.
 ; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and Polynucle
 ; TITLE OF INVENTION: Same
 ; FILE REFERENCE: LEX-0360-USA
 ; CURRENT APPLICATION NUMBER: US/10/189,971
 ; CURRENT FILING DATE: 2002-07-03
 ; PRIOR APPLICATION NUMBER: US 60/302,949
 ; PRIOR FILING DATE: 2001-07-03
 ; PRIOR APPLICATION NUMBER: US 60/315,634
 ; PRIOR FILING DATE: 2001-08-29
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 20
 ; LENGTH: 1207
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-189-971-20
 Alignment Scores:
 Pred. No.: 0.0179 Length: 1207
 Score: 131.50 Matches: 135
 Percent Similarity: 28.72% Conservative: 56
 Best Local Similarity: 20.30% Mismatches: 215
 Query Match: 3.28% Indels: 260
 Ds: 15 Gaps: 36
 US-09-743-237-1 (1-2241) x US-10-189-971-20 (1-1207)
 QY 558 GCCCTCTGGCT-----CAGGAATCCTCTGTC----- 584
 Db 320 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuYr 339
 QY 585 -----AAGTTCCCATCATCCCGAGGAGCA----- 608
 Db 340 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 359
 QY 609 -----GAGGAGCGCTCCAGCTGCCCTCGAAGAAAGACTCCAGCCCATGTGTG 656
 Db 360 CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 379
 QY 657 ATTTCTCAGCTGAAAGGAGGCCGCCAGATGCTCTGCATAGACAACCTGTGGCGCGAGGAG 716
 Db 380 PropheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 395
 QY 717 CTCAAAGCGCTCCATCTCTCTCAGTACGATGACAGACAGTTCCTCCCTCAGTCAGAG 776
 Db 396 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 411
 QY 777 CTCCTTAAGCAATGACAACCTTGTAGTGGAGACACTCTGCCAGTACCAGCAGGAGTTAAT 836
 Db 412 GluProCysAsnLeuCysThrCysLeuGlyGlyPhe----- 423
 QY 837 CTCATCACACAGGTGATAATGGAGCTCTCCCATCAGCTGTCAATGGGCTGCCCTTCCC 896
 Db 424 -----ValThrCysGlyArgArgProCysGluProProGlyCysSerHisPro 439


```

QY 897 TCTGGACCTGCTCTGCAAGGGCCACCACAAATAACTGTCTGCTGGTACTGT----- 947
Db      |||
QY 440 LeuIlePro-----SerGlyHisCysCysProThr 449
QY 948 -----GACTGC-----TTCTCCAGCGGGAC----- 968
Db      |||
QY 450 CysGlnGlyCysArgTyrHisGlyValThrAlaSerGlyLeuThrLeuProAspPro 469
QY 969 -----TTCTGCAACAGCTGCAGCTGCAACAACCTGGCCATCAGCTCGAGCGCTTC 1019
Db      |||
QY 470 LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln----- 485
QY 1020 AAAGCCATAAAGCGCTCTCTGTAGAAATCTTGAAGCTTCCAAACCAAAATGGGAAA 1079
Db      |||
QY 486 -----AspGlyGluGluPheGluGlyProAlaGlySer 496
QY 1080 -----GGCGTCTGGGAGCTGCTAACTTCGA----- 1106
Db      |||
QY 497 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 516
QY 1107 -----CACAGCAAGGGTGCAACTGTAAAGCGC---TCAGGCTGC 1142
Db      |||
QY 517 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 536
QY 1143 CTGAAG----- 1148
Db      |||
QY 537 LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys 556
QY 1149 AACTACTGTGAGTGCATGAG-----GCCAAATCATGTGTTCTTCCATT 1193
Db      |||
QY 557 SerSerCysValCysHisGluGlyValThrCysAlaArgIleGlnCysIleSerSer 576
QY 1194 TGCAAA-----TGCATTGCTTGCACAAACTATGAAGA 1226
Db      |||
QY 577 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 596
QY 1227 AGTCCAGAGCAAAATGTGTATGATGAGCACACCCCTACTACATGAGGCTGGGGAC---TTT 1283
Db      |||
QY 597 GluGlyArgLys-----TyrGluProGlyGluSerPhe 607
QY 1284 GAGAGAGCCATTATTGTGTCGCCAGCAAGTTC-----TCAGGA 1322
Db      |||
QY 608 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 625
QY 1323 CTCCAAAACGTGAGAAAATAGCAGGCGCTTCTCTGTATCTCTGGAAGTAGTGAG 1382
Db      |||
QY 626 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 641
QY 1383 GCCACATGTGCTGCTGCTGGCCAGGGTGAGGAAGCAGCAGGAGCAGCTGTCCCCA 1442
Db      |||
QY 642 CysProProSerGlnLeuLeuProProGly-----ProGlnHisCysCysPro 657
QY 1443 AGCTTGGCTGAGCAGATGATCC----- 1464
Db      |||
QY 658 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 677
QY 1465 -----TGCAGGAGTTTGGCAAGTGC 1484
Db      |||
QY 677 OProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 697
QY 1485 CTGTCCGAGATTCTCCACATCGAGTTCA-----AGTCCAAAGGGCTGAAATTTAG 1535
Db      |||
QY 697 aCysProGluLeuSerCysProLeuSerGluArgHisThrProProGly----- 713
QY 1536 TAGCGTGAAGCTGGTAAGGGGAATGCTGTGGCAGCCTCAGCCCTGGGAATCTGCAC 1595
Db      |||
QY 714 -SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgValAlaAs 733
QY 1596 CGAGGAAGCTGGTGGCCAGGGAGA-----GCAGAGGCCCGCATCATGGCCA 1643
Db      |||
QY 733 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyH1 753

```

```

QY 1644 GTCAAGCTGTGA-----GGTCTGAGTGTCTGCTGCTACTGTCAGCCAGCC 1687
Db      |||
QY 753 sValGluCysHisLeuGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 773
QY 1688 TACTCAAGGTATCTTAAAGTGCAGCAGCAGCAGCCACCCTGG----- 1730
Db      |||
QY 773 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisG1 793
QY 1731 -----GGATGGACACTG----- 1742
Db      |||
QY 793 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 813
QY 1743 -----GCCCTCTCTCTCCCTGGGGAGGCC 1765
Db      |||
QY 813 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 831
QY 1766 CTCCTGGGACTCCTCCCTGCATATAAAGAGAGGCTGATTTCTACTTGTGTGTATGTGTT 1825
Db      |||
QY 832 ---ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 845
QY 1826 TGCTTTCAAAATTCCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCAGCCTCAAGTTAG 1885
Db      |||
QY 846 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 863
QY 1886 AGAGCTAGGCTCT-----TCTTCAGGTGGACTCTGCCAAATCACATACAAGTCAG 1936
Db      |||
QY 863 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 883
QY 1937 GTGGCCATCAGGGGTTTTTCCAGCCAGGCTGTGACAGAGATATATGGAGGGGGTCGG 1996
Db      |||
QY 883 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 902
QY 1997 GTTAGAGCTGG 2007
Db      |||
QY 902 lyValAlaTrp 905

RESULT 6
US-10-189-971-8
; Sequence 8, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-8

Alignment Scores:
Pred. No.: 0.0191 Length: 1477
Score: 131.50 Matches: 135
Percent Similarity: 28.72% Conservativeness: 56
Best Local Similarity: 20.30% Mismatches: 215
Query Match: 3.28% Indels: 260
DB: 15 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-8 (1-1477)
QY 558 GCCCTCTGTGCT-----CAGGAATCTCTGTTC----- 584
Db      |||||

```


845	LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys	864
1143	CTGAAG----- 	1148
865	LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys	884
1149	AAC TACTGTGAGTGC TATGAG----- ::: :::	1193
885	SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer	904
1194	TGCAAA----- :::	1226
905	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis	924
1227	AGTCCAGAAGCAGAAAATGCTCATGAGCACACCCCACTACATGAGAGCTGGGAC- ::: :::	1283
925	GluGlyArgLys----- ::: :::	935
1284	GAGAGCAGCCATATTATTGTCCTCCAGCCCAAGTTC----- :::	1322
936	GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly	953
1323	CTCCAAAACTGAGAAAAATAGCAGGCCTTCTCCTGATCTCCTCGGAAGTAGTGGAG	1382
954	ProProSerLeuArgCysHisArgArg-----GlnCysProSer----- 	969
1383	GCCACATGTGCTGCTGCTGGCCAGGGTGAGGAAGCAGAGGAGCACTGTTCGCCA	1442
970	CysProProSerGlnLeuLeuProGly----- :::	985
1443	AGCTTGCTGAGCAGATGATCC----- :::	1464
986	ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr	1005
1465	----- :::	1484
1005	oProAspProCysTyThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl	1025
1485	CTGTGCGCAGATTCTCCACATCGAGTTCA----- 	1535
1025	acysProGluLeuSerCysProLeuSerGluArgHisThrProGly-----	1041
1536	TAGCGTGCAAGCTGGTTAAAGGGGAATCCCTGTGGCAAGCCTCAGCCCTGGGAATCTGCAC	1595
1042	-SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs	1061
1596	CGAGGAGCTGGTGGCCCGAGGAGGA----- 	1643
1061	pglyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHi	1081
1644	GGTCAGCTGTGA----- 	1687
1081	sValGluCysHisLeuGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVal	1101
1688	TACTCAAGGTATCCTAAAGTGCAAGCAGGAGCCACCCTGG----- 	1730
1101	lProGlnAlaAspSerCysGluArgCysGlnAlaProThrGlnSerCysValHisGln	1121
1731	----- 	1742
1121	nglyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy	1141
1743	----- :::	1765
1141	sMetalGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly-----	1159
1766	CTCTGGGGAGCTCCCTGCGCCATCAAAAAAGAGGTGATTTCTACTTGTGTGTGTGT	1825
1160	-----ProAspLysAlaProAlaLeuSerProGly----- 	1173
1826	TGCTTTCAAAATGCTTAGTAGTACCTCCATTCAGTTATTATTATGAGCCAGCTCAAGTTAG	1885
1174	-----ArgCysLeuProArgProAlaSerCysMetalAlaPheGlyAspProHisLysTrp	1191

QY 1886 AGAGCTAGGCTCT-----TCTTCAGCTGAGCTCTGCCAAATACATACATAAGTCAG 1936
| | | | |
Db 1191 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1211
| | | | |
QY 1937 GTGGCCATCAGGGGTTTCCAGGCCAGGCTGTGACAGAGATATGGAGGGGGTGG 1996
| | | | |
Db 1211 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerg 1230
| | | | |
QY 1997 GTTAGAGCTGG 2007
| | | | |
Db 1230 lyValAlaTrp 1233
| | | | |
RESULT 8
US-10-189-971-4
; Sequence 4, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; TITLE OF INVENTION: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-4

Alignment Scores:
Pred. No.: 0.0196 Length: 1593
Score: 131.50 Matches: 135
Percent Similarity: 28.72% Conservative: 56
Best Local Similarity: 20.30% Mismatches: 215
Query Match: 3.28% Indels: 260
DB: 15 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-4 (1-1593)

QY 558 GCCCTCTGGCT-----CAGGAATCCTGTTCG----- 584
| | | | |
Db 706 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 725
| | | | |
QY 585 -----AAGTTCCTCCATCATCCAGGAGCA----- 608
| | | | |
Db 726 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 745
| | | | |
QY 609 -----GAGGAGGCTCCAGCTGCCCTCGGAAGAAGACTCCAGCCCATGGTG 656
| | | | |
Db 746 CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 765
| | | | |
QY 657 ATTTGTCAGCTGAAGAGGCGCCAGAGCTCTGCATAGACAACTGTGGCGGAGGGAG 716
| | | | |
Db 766 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 781
| | | | |
QY 717 CTCAAGCGCTCATCTGCTCTCCTCAGTACGATGACGAGCAGTTCCTCCTCAGTCAGAG 776
| | | | |
Db 782 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 797
| | | | |
QY 777 CTCCTTAAGCCAATGACAACCTTTAGTGGGAAGACTTCTGCCTAGTACCGAAGTTAAAT 836
| | | | |
Db 798 GluProCysAsnLeuCysThrCysLeuGlyGlyPhe----- 809
| | | | |
QY 837 CTCATCACAGAGTTGATATATGAGAGCTCTCCCATCAGTGTCAATGGGCTGCCTTCCC 896
| | | | |

Db 810 -----ValThrCysGlyArgProCysGluProGlyCysSerHisPro 825
| | | | |
QY 897 TCTGGACCTGCTCTGCAAGGCCACCAATAACTCTGTCTGGTACTGT----- 947
| | | | |
Db 826 LeuIlePro-----SerGlyHisCysCysProThr 835
| | | | |
QY 948 -----GACTGC-----TTCTCCAGCGGGAC----- 968
| | | | |
Db 836 CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspPro 855
| | | | |
QY 969 -----TTCTGCAACAGCTGCAGCTGCAACAACCTGGCCATAGCTCGAGCCCTTC 1019
| | | | |
Db 856 LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln----- 871
| | | | |
QY 1020 AAAGCCATAAAGCGGTCTTGATAGAAATCCTGAAGCTTTCCAAACCAAAATGGGAAA 1079
| | | | |
Db 872 -----AspGlyGluGluPheGluGlyProAlaGlySer 882
| | | | |
QY 1080 -----GCCCTCTCTGGAGCTGCTAAACTCGA----- 1106
| | | | |
Db 883 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 902
| | | | |
QY 1107 -----CACACCAAGGTGCAACTGTAAAGCCG---TCAGGCTGC 1142
| | | | |
Db 903 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 922
| | | | |
QY 1143 CTGAAG----- 1148
| | | | |
Db 923 LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys 942
| | | | |
QY 1149 AACTACTGTGAGTGTATGAG-----GCCAAATCATGTGTTCCTCCATT 1193
| | | | |
Db 943 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 962
| | | | |
QY 1194 TGCAAA-----TGCAATTGCTTGCAGAACTATGAAGA 1226
| | | | |
Db 963 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 982
| | | | |
QY 1227 AGTCCAGAAACGAAATAATGCTGATGAGCACACCCACTACATGAGCTGGGAC---TTT 1283
| | | | |
Db 983 GluGlyArgLys-----TyrGluProGlyGluSerPhe 993
| | | | |
QY 1284 GAGAGCAGCCATTATTTGTCCCGAGCCAAAGTTC-----TCAGGA 1322
| | | | |
Db 994 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 1011
| | | | |
QY 1323 CCTCCAAACTGAGAAAAAATAGGCAGGCTTCTCTGTATCTCTCTGGGAAGTACTGGAG 1382
| | | | |
Db 1012 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 1027
| | | | |
QY 1383 GCCACATGTGCTGCTGCTGCTGCCAGGTGAGGAAGCAGAGCAGCAGCACTGTCCCA 1442
| | | | |
Db 1028 CysProProSerGlnLeuLeuProGly-----ProGlnHisCysCysPro 1043
| | | | |
QY 1443 AGCTGGCTGAGCAGATGATCC----- 1464
| | | | |
Db 1044 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 1063
| | | | |
QY 1465 -----TCGAGAGAGTTTGAAGGTGC 1484
| | | | |
Db 1063 oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 1083
| | | | |
QY 1485 CTGTCCAGATCTCCACATCCAGTTCA-----AGTCCAGGGGCTGAAAATTGAG 1535
| | | | |
Db 1083 aCysProGluLeuSerCysProLeuSerGluArgHisThrProProGly----- 1099
| | | | |
QY 1536 TAGCGTGAAGCTGGTAAAGGGGAATGCTGTGGCAAGCTCAGCCCTGGGAATCTGCAC 1595
| | | | |
Db 1100 -SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgValAlaAs 1119
| | | | |
QY 1596 CGAGGAGAGCTGTGTCGCCAGGAGGA-----GCAGAGGCCCGCATCATGCGCA 1643
| | | | |

```
Db 1119 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHI 1139
QY 1644 GGTCACTGTGA-----GGTCTAGTGATCTGCATGGTACTGCCAGCC 1687
Db 1139 sValGluCysHisLeuGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 1159
QY 1688 TACTCAAGGTATCCTAAAGTCAAGCAGCAGAGCCACCTGG----- 1730
Db 1159 lProGlnAlaAspSerCysGluArgCysGlnAlaProThrGlnSerCysValHisG1 1179
QY 1731 -----GGATGGACACTG----- 1742
Db 1179 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 1199
QY 1743 -----GCCCTCCTGCTCCCTGGGGAGGCC 1765
Db 1199 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 1217
QY 1766 CTCCTGGGACTCCTCCCTGCCTGATATAAAGAGGGTGATTTCTACCTGTGTGTATGTGTT 1825
Db 1218 ----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 1231
QY 1826 TGCTTTCAAATCTTAGTAGTACCTCCATTCAAGTTATATATGAGCCAGCCTCAAGTTAG 1885
Db 1232 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 1249
QY 1886 AGAGCTAGGCTCT-----TCTTCAGGTGACTCTGCCAAATCACATACAAGTCAG 1936
Db 1249 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1269
QY 1937 GTGGCCATCAGGGGTTTTCCAGCCAGGCTCTGACAGAGATATGGAGGGGGTGG 1996
Db 1269 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspArgGlyArgSerg 1288
QY 1997 GTTAGAGCTGG 2007
Db 1288 lyValAlaTrp 1291

RESULT 9
US-10-213-509-5
; Sequence 5, Application US/10213509
; Publication No. US20030054485A1
; GENERAL INFORMATION:
; APPLICANT: Weiss, Joseph
; APPLICANT: Scott, Matthew
; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES
; FILE REFERENCE: STAN-232
; CURRENT APPLICATION NUMBER: US/10/213,509
; PRIOR FILING DATE: 2002-08-06
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4123
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-213-509-5

Alignment Scores:
Pred. No.: 0.0269 Length: 4123
Score: 131.50 Matches: 139
Percent Similarity: 27.79% Conservative: 60
Best Local Similarity: 19.41% Mismatches: 217
Query Match: 3.28% Indels: 300
DB: 15 Gaps: 35

US-09-743-237-1 (1-2241) x US-10-213-509-5 (1-4123)
QY 468 CAAGGAGCAAGCAAGCAAGAAAGTGTGTAATCAAGAACGACGAGTGTGTGTGCA 527
Db 3034 GlnGlyProThrMetGluAlaGluPheCysSerLeuArgProCysProGlyProValPro 3053
```

```
QY 528 GCGCGAGCCCTGAAGACGACGCTTTCCAGGCCCTCTGGCTCAGGAATCCTGTTCAG 587
Db 3054 GlyMetCysProArgAspLysGlnTrpLeu-----AspCysAlaGln 3067
QY 588 TTCCCATCATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 647
Db 3068 GlyProAlaSer----- 3071
QY 648 CCATGGTGTATTGTTCAGCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 707
Db 3071 ----- 3071
QY 708 GCGAGGAGCTCAAGCGCTCCATCTCTTCAGTACGATGACGACGACGACGATTTCCCT 767
Db 3072 CysAlaGluLeuSerAla-----ProArgGlyThrAsnGlnThrCysHisPro 3087
QY 768 CAGTCAGAGCTCCTAAGCCAATGACAACTTTAGTG-----GGAAGACTTTCGCCA 818
Db 3088 GlyCysHisCysProSerGlyMetLeuLeuValSerProArgGlyHisProGlyPro 3107
QY 819 GTACCAAGCAAGTTAAAT-----CTCATCACACAGCTTGATATGAGGCTCTC 866
Db 3108 LeuGlyAlaSerValGlnProProValAlaLeuProGlyAlaIleGlyThrGlySerVal 3127
QY 867 CCATCAGCTGTCAATGGGCTGCTTTCCCTCTGGACCT----- 905
Db 3128 ProGlyAla-----GlyGlyTrpGlyProTrpGlyProTrpSerHisCysSerArgSer 3145
QY 906 -----GCTCTGCAAGGGCCACCCAAAATAACTCTG 935
Db 3146 CysGlyGlyGlyLeuArgSerArgThrArgAlaCysAspGlnProProGlnGlyLeu 3165
QY 936 TCTGGGTACTGTGAC---TGCTTCTCCAGGGGAGCTTCTGCAACAGC----- 980
Db 3166 GlyAspTyrCysGluGlyProArgAlaGlnGlyValCysGlnAlaLeuProCysPro 3185
QY 981 -----TGC--- 983
Db 3186 ValThrAsnCysThrAlaIleGluGlyAlaGluTyrSerProCysGlyProProCysPro 3205
QY 984 ---AGCTGCAACAACCTGGCCGATGAGCTCGAGCGCTTCAAGCCATAAAGGGGTCTTT 1040
Db 3206 ArgSerCysAspAspLeuValHis-----CysVal 3215
QY 1041 GATAGA---AATCTGAAGCTTCCAAACCAAAA----- 1070
Db 3216 TrpArgCysGlnProGlyCysTyrCysProProGlyGlnValLeuSerSerAsnGlyAla 3235
QY 1071 -----ATGGGAAAGGCGCTCTGGAGCTGCTAAACTT-----CGACACAGC 1112
Db 3236 IleCysValGlnProGlyHisCysSerCysLeuAspLeuThrGlyGlnArgHisHis 3255
QY 1113 AAAGGGTCAACTGTAAGCGC---TCAGGCTGCTGAGAACTACTGTAGTCTATGAG 1169
Db 3256 ProGlyAlaArgLeuAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGlu 3273
QY 1170 GCCAAAATCATGTCTTCTCCATTGCAATGCTTCTGCAAAAACTATGAAGAAAGT 1229
Db 3274 GlyArgLeuAsnCysThrAspLeu---ProCysProAspCys----- 3286
QY 1230 CCAGAAGAAATGTGTATGATGACACACCCCACTACATGAGGAGCTGGGGACTTTTGAGAGC 1289
Db 3287 -----GlyGlyGlyGlnSer 3291
QY 1290 AGCCATTATTGTCCCCCAGCAAGTTCTCAGGACCTCCAAACTGAGAAAAATAGGCAG 1349
Db 3292 LeuHis-ProCysGlyGlnProCysProArgSerCysGlnAsp----- 3305
QY 1350 GCCTTCTCTGTATCTCTGGGAAGTAGTGAGGCGCAATGCTGCTGCTGCTGCTGCTGCTGCT 1407
Db 3306 -----LeuSerProGlySerValCysGlnProGlySerValGlyCysGlnProTh 3322
QY 1407 ----- 1407
```


Qy	819	-----GTACCAAGCG-----	-----AAGTTAAATCT	838
		::: :::	::: :::	
Db	252	rHisProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgIyrHisG1	272	
Qy	839	CATCACACAGGTTGATAATGGAGCTCCCATCCAGCTGCTCAATGGGCTGCCCTT	893	
Db	272	yValThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeuCy	292	
Qy	894	-----CCCTCTGGACCTGCTCTGCAAGG	916	
Db	292	sThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLeuCy	312	
Qy	917	GCCACCCAAATAACTCTGTCTGGTACTGTGACTGCTTCTCCAGCGGGGACTTCGCAA	976	
Db	312	oHisProSer-----ProGlyProCysPheCys-----ProValCysH1	325	
Qy	977	CAGCTGCAGCTGCACAACCTGCGCCATGAGCTCGAGCGCTTCAAGCCATAAAGCGCTG	1036	
Db	325	sSerCysLeuSerGlnGlyArgGluHisGln-----	335	
Qy	1037	TCTTGATAGAAATCTCAAGCTTTCACCAAAAAATGGGAAA-----	1079	
Db	336	-----AspGlyGluGluPheGluGlyProAlaGlySerCysGluTrpCysArgCy	352	
Qy	1080	-----GGCGTCTGGAGCTGCTAACTTCGA-----	1106	
Db	352	sGlnAlaGlyGlnValSerCysValArgLeuGlnCysProProLeuProCysLysLeuG1	372	
Qy	1107	-----CACAGCAAAAGGTGCAACTGTAGCGC---TCAGGCTGGCTGAAG-----	1148	
Db	372	nValThrGluArgGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluG1	392	
Qy	1149	-----AACTACTGTGAGTGCTA	1165	
Db	392	uHisProGluGlySerArgTrpValProProAspSerAlaCysSerSerCysValCysH1	412	
Qy	1166	-----GCCAAATCATGTGTTCTTCCTATTTGCGAAA-----	1199	
Db	412	sGluGlyValValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnProArgG1	432	
Qy	1200	-----TGCATTGCTTCGCAAAACTATGAAGAAGTCCAGCAACGAAAAT	1243	
Db	432	nGlyProHisAspCysCysProGlnCysSerAspCysGluHisGluGlyArgLys-----	450	
Qy	1244	GCTGATGAGCACACCCACTACATGAGCGCTGGGAGC---TTTGAGAGCAGCCATATTT	1300	
Db	451	-----TyrGluProGlyGluSerPheGlnProGly-----Al	461	
Qy	1301	GTCCCCAGCAAGTTC-----TCAGGACCTCCAAAACCTGAGAAA	1339	
Db	461	aAspProCysGluValCysIleCysGluProGlnProGluGlyProProSerLeuArgCy	481	
Qy	1340	AAATAGCAGCGCTTCTCTGTATCTCTCTGGGAAGTAGTGGAGCCACATGTGCTGCGCT	1399	
Db	481	sHisArgArg-----GlnCysProSer-----LeuValGlyCysProProSerGlnLe	497	
Qy	1400	GCTGGCCCGGTTGAGGAACAGCAGCAGGAGCACTGTTCCCAAGCTTGCTGAGCAGAT	1459	
Db	497	uLeuProProGly-----ProGlnHisCysCysProThrCysAlaGluAlaLe	513	
Qy	1460	GATCC-----	1464	
Db	513	u-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrT	533	
Qy	1465	-----TGGAGGAGTTTGGAAAGTGGCTGTGCGAGATTTCTCCA	1501	
Db	533	hrcysGlnCysGlnAspLeuThrTrpLeuCystIleHisGlnAlaCysProGluLeuSerC	553	
Qy	1502	CATCGAGTTCA-----AGTCCAAGGGCTGAAAATTTAGTAGGTCTCAAGCTGGTA	1552	
Db	553	ysProLeuSerGluArgHisThrProProGly-----SerCysCysProValC	569	
Qy	1553	AAGGGGAATGCCTGTGGCAAGCCTTCAGCCCTGGGAAATCTCACCCGGAAGACTGGTGCC	1612	

```

Db      569 ysArgGluCysValValGluAlaGluCluArgArgValAlaAAspGlyGluSerTrpArgA 589
Qy      1613 AGGAGGA-----GCAGAGCCCGCATCATGCCACGGTCAGCTGTGA-----1655
Db      589 spProSerAsnAlaCysIleAlaCysThrCysHlsArgGlyHlsValGluCysHisLeuG 609
Qy      1656 -----GGTCTGAGTGATCTGCATCGTACTGCCACGCCCTACTCAAGGTATCCTAA 1704
Db      609 luGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysValProGlnAlaAspSerc 629
Qy      1705 AGTCAACAGCAGGAGCACCCCTGG-----1730
Db      629 yScysGluArgCysGlnAlaProThrGlnSerCysValHisGlnGlyArgGluValAlaAs 649
Qy      1731 -----GATGCACACTG-----1742
Db      649 erGlyGluArgTrpThrValAspThrCysThrSerCysSerCysMetalaglyThrValA 669
Qy      1743 -----GCCTCTGCTGCCCTGGGAGGCCCTCTGGGACCTCCCTGC 1782
Db      669 rgCysGlnSerGlnArgCysSerProLeuSerCysGly-----ProAsPlysAlap 686
Qy      1783 CTTGCATAAAAAGAGGGTGATTTCTACTTGTGTGTATTGTTGTTTCAAATTCCTTA 1842
Db      686 roAlaLeuSerProGly-----SerCysCysPro-----ArgCysLeuP 699
Qy      1843 GTAGTACCCTCAITCAAGTTATTATGAGCAGCGCTCAAGTTAGAGAGCTAGGCTCT---1898
Db      699 roArgProAlaSerCysMetalapheGlyAspProHisTyArgThrPheAspGlyArgL 719
Qy      1899 -----TCTCAGGTGGTACTTGCCCAANTCACATACAAGTCAGTGCCCATCAGGGGTTT 1953
Db      719 euLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysAspCysHisSerGlyAsp 738
Qy      1954 TTCAGGCCAGGCTGTGACAGAGAGATATCGGAGGGGGTTCGGGTTAGAGCTGG 2007
Db      739 PheSerValHis---ValThrAsnAspAspArgGlyArgSerglyValAlatrp 755

RESULT l1
US-10-189-971-16
; Sequence 16: Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Poly
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1251
; TYPE: PR1
; ORGANISM: homo sapiens
US-10-189-971-16

Alignment Scores:
Pred. No.: 0.0804 Length: 1251
Score: 124.00 Matches: 135
Percent Similarity: 28.32% Conservative: 59
Best Local Similarity: 19.71% Mismatches: 227
Query Match: 3.09% Indels: 265
DB: 15 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-16 (1-1251)
```

US-09-743-237-1 (1-2241) x US-10-189-971-16 (1-1251)

Qy	558	GCCCTCTGGCT	-----CAGGAATCTCTGTGC-----	584
Db	329	AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr	-----AAGTCTCCATCATCCAGGAGCA-----	608
Qy	585	-----		
Db	349	GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu	-----GAGGAGGCGCTCAGCTGCCTCGGAAGAAGAACTCCAGCCCATGGT	656
Qy	609	-----		
Db	369	CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys	-----CysCysProAspCysAspGlyCysGlu	404
Qy	657	ATTGTGTCAGCTGAAGAGGCGCCAGATGCTGTGCATAGACAACACTGCGCGGAGGGAG	-----GTACACGCGCTCCATCTGCTCAGTACGATCACCAGACGAGCATTTCCCTCAGTACGAG	776
Db	389	ProPheProAlaArgGlyAsp-----	-----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg	420
Qy	717	CTCAAAACCGCTCCATCTGCTTCCTCAGTACGATCACCAGACGAGCATTTCCCTCAGTACGAG	-----GTACACGCGCTCCATCTGCTCAGTACGATCACCAGACGAGCATTTCCCTCAGTACGAG	776
Db	405	-----	-----	
Qy	777	CTCCTAAGCAATGACAACACTTGTAGTGGGAAAGACTTCTGCCA	-----	818
Db	421	GluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCysGlyArgArgProCys	-----	440
Qy	819	-----GTACACGCG-----	-----	827
Db	441	GluProProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGln	-----	460
Qy	828	-----AAGTTAAATCTCATCACACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCAT	-----	881
Db	461	GlyCysArgTyrHisGlyValThrAlaSerGlyGluThrLeuProAspProLeuAsp	-----	480
Qy	882	GGGCGTGCCTTT-----	-----CCCTCT	899
Db	481	ProThrCysSerLeuCysThrCysGlnGluGlySerMetArgCysGlnLysLysProCys	-----	500
Qy	900	GGACCTGCTCTGCAGGGCCACCCAAATAACTCTGTCTGGTACTGTACTGCTTCTCC	-----	959
Db	501	AlaProAlaLeuCysProHisProSer-----	-----ProGlyProCysPheCys-----	515
Qy	960	AGCGGGAGCTTCGCAACAGCTGCAGCTGCAACCAACCTGCGCCATGAGCTCGAGCGGCTC	-----	1019
Db	516	-----ProValCysHisSerCysLeuSerGlnGlyArgGluHisGln-----	-----	529
Qy	1020	AAAGCCATAAGGCGTCTGTGATAGAAATCCTGAAGCTTTCCCAACCAAAATGGGAAA	-----	1079
Db	530	-----	-----AspGlyGluGluPheGluGlyProAlaGlySer	540
Qy	1080	-----GGCGTCTGGGAGCTGCTAACTTCA-----	-----	1106
Db	541	CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro	-----	560
Qy	1107	-----	-----CACAGCAAAAGGGTGCAACTTGAGCC-----TCAGCGCTGC	1142
Db	561	LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys	-----	580
Qy	1143	CTGAAG-----	-----	1148
Db	581	LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys	-----	600
Qy	1149	AACTACTGTGAGTCTATGAG-----	-----GCCAAATCATGTCTTCTTCCATT	1193
Db	601	SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer	-----	620
Qy	1194	TGCAAA-----	-----TGCAATTCCTGCAAAAACTATGAGAA	1226
Db	621	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis	-----	640
Qy	1227	AGTCCAGACGAAAAATGCTGATGAGCACACCCCATACATGAGCTGGGAG-----	-----TTT	1283
Db	641	GluGlyArgLys-----	-----TyrGluProGlyGluSerPhe	651

RESULT 12
US-10-189-971-24
; Sequence 24, Application US/10189971
; Publication NO. US20030028907A1
; GENERAL INFORMATION:

APPLICANT: Walke, D. Wade
APPLICANT: Scoville, John
TITLE OF INVENTION: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
FILE REFERENCE: Lex-0360-USA
CURRENT APPLICATION NUMBER: US/10/189, 971
CURRENT FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/302, 949
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: US 60/315, 634
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 1342
TYPE: PRT
ORGANISM: homo sapiens
US-10-189-971-24

Alignment Scores:
Pred. No.: 0.0823 Length: 1342
Score: 124.00 Matches: 135
Percent Similarity: 28.32% Conservative: 59
Best Local Similarity: 19.71% Mismatches: 227
Query Match: 3.09% Indels: 265
DB: 15 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-24 (1-1342)

QY	558	GCCCCCTGTGGCT	-----CAGGAATCCTGTGGC-----	584
Db	420	AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr		439
QY	585	-----AAGTCCCATCATCCAGGAGCA-----		608
Db	440	GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu		459
QY	609	-----GAGGAGGCTCCAGCTGCCTCGGAAGAAAGACTCCAGCCCATGGTG		656
Db	460	CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys		479
QY	657	ATTGTGACGTGAAGGAGCGCCAGATGCTCTGCATAGACAACTGTGGCGGAGGAG		716
Db	480	ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu		495
QY	717	CTCAAGCGCTCATCTGCTTCCTCAGTAGCAGCAGGAGTTCCTCAGTCAGAG		776
Db	496	-----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg		511
QY	777	CTCCCTAAGCAATGACAACCTTAGTGGGAAGACTTCTGCCA-----		818
Db	512	GluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCysGlyArgArgProCys		531
QY	819	-----GTACCAGCG-----		827
Db	532	GluProProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGln		551
QY	828	-----AAGTTAATCTCATCACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCAT		881
Db	552	GlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspProLeuAsp		571
QY	882	GGGGTGCCTTT-----		899
Db	572	ProThrCysSerLeuCysThrCysGlnGlySerMetArgCysGlnLysLysProCys		591
QY	900	GGACCTGCTCTGAAGGGCCACCCAAATAACTCTGTCTGGTACTGTGACTGCTTCTCC		959
Db	592	AlaProAlaLeuCysProHisProSer-----ProGlyProCysPheCys-----		606
QY	960	AGCGGGGACTTCTGCACACAGCTGAGCTGCACAACTGCGCCATGAGCTCGAGCGCTTC		1019
Db	607	-----ProValCysHisSerCysLeuSerGlnGlyArgGluHisGln-----		620

QY	1020	AAAGCCATAAAGCGCTCTCTTGATAGAAATCTGAAGCTTTTCAACCAAAATGGGAAA	1079
Db	621	-----AspGlyGluGluPheGluGlyProAlaGlySer	631
QY	1080	-----GCCCGTCTGGAGCTGCTAAACTTCA-----	1106
Db	632	CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro	651
QY	1107	-----CACAGCAAAAGGTGCAACTGTAAAGCC-----TCAGGCTGC	1142
Db	652	LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys	671
QY	1143	CTGAAG-----	1148
Db	672	LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys	691
QY	1149	AACTACTGTGAGTGCATGAG-----GCCAAAATCATGTCTTCTTCCATT	1193
Db	692	SerSerCysValCysHisGlyValValThrCysAlaArgIleGlnCysIleSerSer	711
QY	1194	TGCAAA-----TGCATTCTTCCAAAATATGAGAA	1226
Db	712	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis	731
QY	1227	AGTCCAGACGAAAAATGCTGATGAGCACACCCACTACATGAGAGCTGGGAC-----TTT	1283
Db	732	GluGlyArgLys-----TyrGluProGlyGluSerPhe	742
QY	1284	GAGAGAGCCATTATTGTGCCAGCAAGTTC-----TCAGGA	1322
Db	743	GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly	760
QY	1323	CTCCAAAATGAGAAAAATAGCAGGCCCTTCTCTGTATCTCTCTGGAGTAGTCGAG	1382
Db	761	ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly	776
QY	1383	GCCACATGTCCTGCTGCTGGCCAGGCTGAGGAGCAGCAGCAGTGTGCCCA	1442
Db	777	CysProProSerGlnLeuLeuProGly-----ProGlnHisCysCysPro	792
QY	1443	AGTGTGCTCAGCAGATGATCC-----	1464
Db	793	ThrCysAlaGluAlaLeu-SerAsnSerGluGlyLeuLeuGlySerGluLeuAlaPr	812
QY	1465	-----TGGAGGAGTTTGGAGGTGC	1484
Db	812	oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl	832
QY	1485	CTCTCGCAGATTCTCCACATCGAGTTCA-----AGTCCAAAGGGCTGAAATTTGAG	1535
Db	832	aCysProGluLeuSerCysProLeuSerGluArgHisThrProGly-----	848
QY	1536	TAGCTGCAAGCTGGTAAAGGGGAATCCCTGTGGCAAGCCTCAGCCCTGGGAATCTGCAC	1595
Db	849	-SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs	868
QY	1596	CGAGGAGCTGGTCCCGAGGAGGA-----GCAGAGCCCGCATCATGCCCA	1643
Db	868	pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHis	888
QY	1644	GGTCAGCTGTGA-----GGTCTGAGTGTGATCTGCATCTGCTACTGCGCCAGCC	1687
Db	888	sValGluCysHisLeuGluCysGlnAlaLeuSerCysCysProHisGlyTrpAlaLysVa	908
QY	1688	TACTCAAGGTATCTCAAGTGCAGAGCAGCAGCCACCTGG-----	1730
Db	908	lProGlnAlaAspSerCysGluArgCysGlnAlaProThrGlnSerCysValHisG1	928
QY	1731	-----GGATGGACACTG-----	1742
Db	928	nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy	948


```
QY 1465 -----TGCAGGAGTTTGAAGGTGC 1484
Db 982 oProAspProCysTyThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 1002
QY 1485 CTGTCGACAGATTCACATCGAGTTCA-----AGTCCAAAGGGCTGAAATTTGAG 1535
Db 1002 acYsProGluLeuSerCysProLeuSerGluArgHisThrProProGly----- 1018
QY 1536 TAGCGTGAAGCTGGTAAAGGGCAATGCCFTGTGGCAAGCTCAGCCCTGGGAATCTGCAC 1595
Db 1019 -SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs 1038.
QY 1596 CGAGGAAGCTGGTCCCGAGGAGGA-----GCAGAGGCCGCGCATCATGGCCA 1643
Db 1038 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHi 1058
QY 1644 GGTACAGCTGTGA-----GGTCTAGTGTCTCGATGGTACTGGCCAGCC 1687
Db 1058 sValGluCysHisLeuGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 1078
QY 1688 TACTCAAGGTATCTTAAGTGCAAGCAGCAGCAGCCACCTGG----- 1730
Db 1078 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisGl 1098
QY 1731 -----GGATGGACACTG----- 1742
Db 1098 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 1118
QY 1743 -----GCCCTCCTGCTCCCTGGGGAGCC 1765
Db 1118 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 1136
QY 1766 CTCTGGGAGCTCCCTGCGCTGCATATAAAGAGAGGTGATTTCTACTGTGTGTATGTGT 1825
Db 1137 -----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 1150
QY 1826 TGCTTTCAATTCCTAGTAGTACCTCCATTCAAGTTATTATGAGCCAGCCTCAAGTTAG 1885
Db 1151 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyAr 1168
QY 1886 AGAGCTAGGCTCT-----TCTTCAGTGGACTCTGCCCAATACATACATAGTCAG 1936
Db 1168 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrrValLeuAlaLysA 1188
QY 1937 GTGGCCATCAGGGGTTTTCCAGCCAGCCCTGTGACAGAGATATGGGAGGGGGTCGG 1996
Db 1188 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerg 1207
QY 1997 GTTAGAGCTGG 2007
Db 1207 lyValAlaTrp 1210

RESULT 14
US-10-189-971-12
; Sequence 12, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucleot
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
```

```
; LENGTH: 1570
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-12

Alignment Scores:
Pred. No.: 0.0867 Length: 1570
Score: 124.00 Matches: 135
Percent Similarity: 28.32% Conservative: 59
Best Local Similarity: 19.71% Mismatches: 227
Query Match: 3.09% Indels: 265
DB: 15 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-12 (1-1570)
QY 558 GCCCTCTGGCT-----CAGGAATCCTGTTGC----- 584
Db 648 AlaProCysAlaHisProArgGlnGlyProCysProSerCysAspGlyCysLeuTyr 667
QY 585 -----AAGTTCCCATCATCCAGGAGGCA----- 608
Db 668 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 687
QY 609 -----CAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTG 656
Db 688 CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 707
QY 657 ATTTGTCAGCTGAAGGAGGCGCCAGATCTCTGCATACACAACTGTGGCGGAGGAG 716
Db 708 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 723
QY 717 CTCAAAGCGCTCCATCTGCTTCCTCAGTACGATGACACAGAGAGTTTCCCTCAGTCAGAG 776
Db 724 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 739
QY 777 CTCCCTAAGCAAAGACAACTTTAGTGGGAAGACTTCTGCCA----- 818
Db 740 GluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCysGlyArgArgProCys 759
QY 819 -----GTACCAGG----- 827
Db 760 GluProProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGln 779
QY 828 -----AAGTTAAATCTCATCACACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCAT 881
Db 780 GlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspProLeuAsp 799
QY 882 GGGGTGCCCTTT----- 899
Db 800 ProThrCysSerLeuCysThrCysGlnGluGlySerMetArgCysGlnLysLysProCys 819
QY 900 GGACCTGCTCTGCAGGGCCACCCAAATAACTCTGTCTGGTACTGTGACTGCTTCTCC 959
Db 820 AlaProAlaLeuCysProHisProSer-----ProGlyProCysPheCys----- 834
QY 960 AGCGGGGACTTCTGCAACAGCTGCAGCTGCAACAACCTGCGCATAGCTCGAGCGCTTC 1019
Db 835 -----ProValCysHisSerCysLeuSerGlnGlyArgGluHisGln----- 848
QY 1020 AAAGCCATAAAGCGCTCTTTGATAGAAATCTCGAAGCTTTCCAAACCAAAATGGGAAA 1079
Db 849 -----AspGlyGluGluPheGluGlyProAlaGlySer 859
QY 1080 -----GGCCGCTCGGAGCTGCTAAACTTCGA----- 1106
Db 860 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 879
QY 1107 -----CACAGCAAGAGGTGCAACTGTGAAGCC-----TCAGGCTGC 1142
Db 880 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 899
QY 1143 CTGAAG----- 1148
Db 1143 CTGAAG-----
```

Db 900 LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys 919
 QY 1149 AACTACTGTGAGTGTATGAG-----GCCAAATCATGTGTCTTCCCAAT 1193
 Db 920 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 939
 QY 1194 TCGAAA-----TGCATTGCTTGCAAAACAACTATGAAGA 1226
 Db 940 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 959
 QY 1227 AGTCCAGACGAAAAATGCTGATGAGCACACCCACTACATGAGCCCTGGGAC---TTT 1283
 Db 960 GluGlyArgLys-----TyrGluProGlyGluSerPhe 970
 QY 1284 GAGAGCAGCCATTATTGTCCTCCAGCCCAAGTTC-----TCAGGA 1322
 Db 971 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 988
 QY 1323 CCTCCAAAACAGAGAAAAATAGGAGCCCTTCTCTGTATCTCTGGAAGTAGTGAG 1382
 Db 989 ProProSerLeuArgCysHisArg-----GlnCysProSer-----LeuValGly 1004
 QY 1383 GCCACATGTGCTGCTGCTGCCCGCCAGGTGAGGAGCAGCAGGAGCAGCACTTCCCA 1442
 Db 1005 CysProProSerGlnLeuLeuProProGly-----ProGlnHisCysCysPro 1020
 QY 1443 AGCTTGCTGAGCAGATGATCC-----1464
 Db 1021 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 1040
 QY 1465 -----TGAGAGGAGTTTGAAGGTGC 1484
 Db 1040 oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 1060
 QY 1485 CTGTCGCACATTCACATCAGATCA-----AGTCCAGGGGCTGAAATTGAG 1535
 Db 1060 acysProGluLeuSerCysProLeuSerGluArgHisThrProProGly-----1076
 QY 1536 TAGCGTCGAAGTGTGTAAGGGGAATCCCTGTGGAAGCCTCAGCCCTGGGAATCTGCAC 1595
 Db 1077 SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs 1096
 QY 1596 CGAGGAAGCTGTGCTCCAGGGAGGA-----GCAGAGGCGCCGCATCATGGCCA 1643
 Db 1096 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHi 1116
 QY 1644 GGTACAGCTGA-----GGTCTGAGTGATCTGCATGTCATGGCCAGCC 1687
 Db 1116 sValGluCysHisLeuGluGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 1136
 QY 1688 TACTCAAGTATCTTAAGTGAAGCAGCAGCAGCCACCTCG-----1730
 Db 1136 lProGlnAlaAspSerCysGluArgCysGlnAlaProThrGlnSerCysValHisGln 1156
 QY 1731 -----GGATGGACACTG-----1742
 Db 1156 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 1176
 QY 1743 -----GCCCTCTCTGCTCCCTGGGAGGCC 1765
 Db 1176 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly-----1194
 QY 1766 CTCTGGGGACTCCCTGCTCCCTGCATAAAAGAGGGTGATTTCTACTGTGTGTATGTGT 1825
 Db 1195 -----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro-----1208
 QY 1826 TCGTTTCAATTGCTTAGTAGTACCTCCATTCAAGTTATATGACCCACCCCTCAAGTTAG 1885
 Db 1209 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 1226
 QY 1886 AGAGCTAGGCTCT-----TCCTTACGTTGACTCTGCCCAAAATCACATACAAATGCTAG 1936
 Db 1226 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1246

QY 1937 GTGGCCATCAGGGTGTTCAGGCCAGGCCTGTGACAGGAGATATGGAGGGGGTCCG 1996
 Db 1246 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 1265
 QY 1997 GTTAGAGCTGG 2007
 Db 1265 LyValAlaTrp 1268
 RESULT 15
 US-10-189-971-2
 ; Sequence 2, Application US/10189971
 ; Publication NO. US20030028907A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Scoville, John
 ; APPLICANT: Turner, C. Alexander Jr.
 ; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle
 ; TITLE OF INVENTION: Same
 ; FILE REFERENCE: LEX-0360-USA
 ; CURRENT APPLICATION NUMBER: US/10/189,971
 ; CURRENT FILING DATE: 2002-07-03
 ; PRIOR APPLICATION NUMBER: US 60/302,949
 ; PRIOR FILING DATE: 2001-07-03
 ; PRIOR APPLICATION NUMBER: US 60/315,634
 ; PRIOR FILING DATE: 2001-08-29
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1628
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-189-971-2
 Alignment Scores:
 Pred. No.: 0.0878 Length: 1628
 Score: 124.00 Matches: 135
 Percent Similarity: 28.32% Conservative: 59
 Best Local Similarity: 19.71% Mismatches: 227
 Query Match: 3.09% Indels: 265
 DB: 15 Gaps: 36
 US-09-743-237-1 (1-2241) x US-10-189-971-2 (1-1628)
 QY 558 GCCCTCTGGCT-----CAGGAATCCTGTTC-----584
 Db 706 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 725
 QY 585 -----AAGTTCCCATCATCCAGGAGCA-----608
 Db 726 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 745
 QY 609 -----GAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGTGTG 656
 Db 746 CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 765
 QY 657 ATTTGTACGTGAAGAGGAGGCCCGCAGATGCTCTGCATGACAACTGTGGCGGAGGAG 716
 Db 766 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 781
 QY 717 CTCGAAGCCCTCCATCTGCTCTCTCAGTAGATGACGAGCAGCTTCCCTCAGTCAGAG 776
 Db 782 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 797
 QY 777 CTCCTTAAGCCCAATGACAACCTTTAGTGGGAAGACTTCTGCCA-----818
 Db 798 GluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCysGlyArgArgProCys 817
 QY 819 -----GTACCAGCG-----827
 Db 818 GluProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGln 837
 QY 828 -----AAGTTAAATCTCATCACAGGTTGATATGGAGCTCTCCCATCAGCTGCAAT 881

```

Db      838 GlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspProLeuAsp 857
QY      882 GGGGCTGCCCTT-----CCTCT 899
Db      858 ProThrCysSerLeuCysThrCysGlnGluGlySerMetArgCysGlnLysLysProCys 877
QY      900 GGACCTGCTCTGCAAGGGCCACCAAAATAACTCTGTCTGGTACTGTGACTCTGCTCTCC 959
Db      878 AlaProAlaLeuCysProHisProSer-----ProGlyProCysPheCys----- 892
QY      960 AGCGGGGACTTTCACACAGCTCGAGCTGCAACAACCTGCGGCATGAGCTCGAGCGCTTC 1019
Db      893 -----ProValCysHisSerCysLeuSerGlnGlyArgGluHisGln----- 906
QY      1020 AAAGCCATAAGCGGTCTTGATAGAAATCCTGAAGCTTTCACCAACCAAAATGGGAAA 1079
Db      907 -----AspGlyGluGluPheGluGlyProAlaGlySer 917
QY      1080 -----GGCCGTCTGGAGCTGCTAAACTTCGA----- 1106
Db      918 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 937
QY      1107 -----CACAGCAAGGTCGAACGTGAAGCGC-----TCAGGCTGC 1142
Db      938 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 957
QY      1143 CTGAAG----- 1148
Db      958 LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProAspSerAlaCys 977
QY      1149 AACTACTGTGAGTGCATGAG-----GCCAAATCATGTCTTCTTCCATT 1193
Db      978 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 997
QY      1194 TGCAAA-----TGCATTGCTTGCACAAAACATGAAGAA 1226
Db      998 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 1017
QY      1227 AGTCCAGAGCAAAATGTGTAGCAGCACACCCCTACATGAGCGCTGGGAC---TTT 1283
Db      1018 GluGlyArgLys-----TyrGluProGlyGluSerPhe 1028
QY      1284 GAGCAGCAGCATTTATTGTCCCGCAGCAAGTTC-----TCAGGA 1322
Db      1029 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 1046
QY      1323 CTCCCAAACTGAGAAAAATAGCAGGCGCTTCTCGTGTATCTCTGGGAAGTAGTGGAG 1382
Db      1047 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 1062
QY      1383 GCCACATGTCCTGCTGCTGGCCCGCAGGAGGAGNACGAGCAGGAGCACTGTTCCCA 1442
Db      1063 CysProProSerGlnLeuLeuProProGly-----ProGlnHisCysCysPro 1078
QY      1443 AGCTTGGCTGACGAGATGATCC----- 1464
Db      1079 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 1098
QY      1465 -----TGGAGGAGTTTGGAGGTGC 1484
Db      1098 oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 1118
QY      1485 CTCTCGCAGATTCTCCACATCGAGTCA-----AGTCCAAAGGGCTGAAAATTGAG 1535
Db      1118 aCysProGluLeuSerCysProLeuSerGluArgHisThrProGly----- 1134
QY      1536 TAGGTGCAAGCTGGTAAAGGGGAATGCCGTGTGGCAAGCCTCAGCCCTGGGAATCTGCAC 1595
Db      1135 -SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs 1154
QY      1596 CGAGGAAGCTGGTGCCTCCAGGGAGGA-----GCAGAGGCCCGCATCATGGCCA 1643

```

Search completed: July 24, 2003, 14:43:23
 Job time : 145.957 secs

```

Db      1154 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHI 1174
QY      1644 GGTCAAGTGA-----GGTCTGAGTGATCTGCATGCTACTGCTACTGCCAGCC 1687
Db      1174 sValGluCysHisLeuGluGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 1194
QY      1688 TACTCAAGGTATCTCTAAAGTGCAAGCAGCAGCCACCCCTGG----- 1730
Db      1194 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrClnSerCysValHisG1 1214
QY      1731 -----GGATGGACACTG----- 1742
Db      1214 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 1234
QY      1743 -----GCCCTCTCTCTCCCTGGGAGGCC 1765
Db      1234 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 1252
QY      1766 CTCTGGGACTCCCTGCTGCATATAAAAGAGGGTGATTTCTACTTGTGTGTATGTGTT 1825
Db      1253 -----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 1266
QY      1826 TGCTTCAAAFTGCTTAGTAGTACCTCCATTCAAGTTTATTATGAGCCAGCCTCAAGTTAG 1885
Db      1267 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 1284
QY      1886 AGAGCTAGGCTCT-----TCTTCAGGTGGACTCTGCCCAATCACATACAGTCAG 1936
Db      1284 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1304
QY      1937 GTGGCCATCAGGGGTTTTTCCAGGCCAGGCTGTGACAGAGAGATATGGAGGGGGTCCG 1996
Db      1304 spCysHisSerGlyAspPheSerValHis-----ValThrAsnAspAspArgGlyArgSerG 1323
QY      1997 GTTAGAGCTGG 2007
Db      1323 lyValAlaTrp 1326

```


Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	143.5	3.6	449	4	US-09-252-991A-23908	Sequence 23908, A
2	125.5	3.1	1184	4	US-09-266-225D-18	Sequence 18, Appl
C 3	120.5	3.0	697	4	US-09-252-991A-24009	Sequence 24009, A
C 4	119.5	3.0	1185	3	US-09-041-886-23	Sequence 23, Appl
C 5	119.5	3.0	2594	3	US-08-718-388-7	Sequence 7, Appl
C 6	119.5	3.0	5405	3	US-08-718-388-9	Sequence 9, Appl
C 7	119	2.9	246	4	US-09-252-991A-25102	Sequence 25102, A
C 8	118.5	3.0	169	4	US-09-252-991A-28109	Sequence 28109, A
9	118.5	3.0	3118	4	US-09-579-181-1	Sequence 1, Appl
10	117.5	2.9	567	4	US-09-252-991A-28481	Sequence 28481, A
11	117.5	2.9	2972	4	US-09-579-181-2	Sequence 2, Appl
C 12	116	2.9	291	4	US-09-252-991A-17495	Sequence 17495, A

Qy	624	AGCTGGAGGCGCTCTCTCGCTCTCTGGGATGGGAACCTTGCAACAGGATTTCTCTGAGCCA	565
Db	659	SerPheArgThrGlyThrProProGlyTyrArgGlyThr-----SerProPro	674
Qy	564	GAGGCGCTGGAAAGCTCGCTCTTCAGGCGCTCCG-----CCTGGCACACTACCAC	514
Db	675	AlaGlyProGlyThrPhelYs--ProGlySerProThrValGlyProGlyProLeuPro	694
Qy	513	CTGCT 509	
Db	694	IOALA 695	
RESULT 5			
US-08-718-388-7			
; Sequence 7, Application US/08718388			
; Patent No. 6271362			
; GENERAL INFORMATION:			
; APPLICANT: MORIKAWA, MINORU			
; APPLICANT: HARADA, NAOKI			
; TITLE OF INVENTION: GENE ENCODING IgG FC REGION-BINDING			
; TITLE OF INVENTION: PROTEIN			
; NUMBER OF SEQUENCES: 29			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH			
; STREET: PO BOX 747			
; CITY: FALLS CHURCH			
; STATE: VA			
; COUNTRY: USA			
; ZIP: 22040-0747			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/718,388			
; FILING DATE:			
; CLASSIFICATION: 536			
; ATTORNEY/AGENT INFORMATION:			
; NAME: MURPHY JR, GERALD M			
; REGISTRATION NUMBER: 28,977			
; REFERENCE/DOCKET NUMBER: 0230-111			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (703) 205-8000			
; TELEFAX: (703) 205-8050			
; INFORMATION FOR SEQ ID NO: 7:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 2594 amino acids			
; TYPE: amino acid			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
US-08-718-388-7			
Alignment Scores:			
Pred. No.: 0.00508 Length: 2594			
Score: 119.50 Matches: 122			
Percent Similarity: 31.83% Conservative: 55			
Best Local Similarity: 21.94% Mismatches: 162			
Query Match: 2.95% Indels: 218			
DB: 3 Gaps: 34			
US-09-743-237-1 (1-2241) x US-08-718-388-7 (1-2594)			
Qy	1479	TTCCAAACTCTCCAGGATCACTGCT-----CAGCCAAGCTTGGGAACACT-----	143
Db	1401	PheGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu	142
Qy	1431	-----GCTCCTGCTCTGCTTCTCACCTCGGCCACGACGC-----	139
Db	1421	GluValValProAspSerProCysLeuProThrProCysProGlySerGluAsp	144
Qy	1395	-----AGGCACATGTGGCCTCCACTACTTCCAGGAGATACAGGAGAACGGCTGCC	134

Qy	973	-----GCAACAGCTCGAGCTGCAACAACTGCGCC	1002
Db	2444	AlaLeuAlaSerProGluSerLeuGluLeuAlaSerValAlaSerSerGluThrSerSer	2463
Qy	1003	ATAGCTCGAGCGCTCAAGCCATAAAGCGGTCTTGATAGAAATCCCTGAAGCTTTC	1062
Db	2464	LeuSerLeuValProProLysAspLeuLeuProValAlaValGluIleLeuProValSer	2483
Qy	1063	AACCAAAATGGGAAAGCGCTGGGAGCTGCTAAACTTCGACACAGCAAGGGTGCA	1122
Db	2484	GluLys-----AsnLeuSerLeuThrProSerAlaProSerLeu	2496
Qy	1123	ACTGTAAGCGCT-----CAGGCTGCCTGA	1146
Db	2497	ThrLeuGluAlaGlySerIleProAsnGlyGlnGluGluAlaProAspSerAlaGlu	2516
Qy	1147	AGAACTACTGTGAGTGTATGAGCCNAATCA-----TGTGTT-----	1185
Db	2517	GlyThrThrLeuThrValLeu-----ProGluGlyGluGluLeuProLeuCysValSerGlu	2535
Qy	1186	-----CTTCCATTGCAATGCATT-----	1205
Db	2536	SerAsnGlyLeuGluLeuProProSerAlaAlaSerAspGluProLeuGluProLeu	2555
Qy	1206	---GCTGTCAAAATATGAAGAAGTCCAGAACGAAAAATGCTGATGAGCACACCCAC	1262
Db	2556	GluAlaAspArgThrSerGluGluLeuThrGluAlaLysThrProThrSerSerProGlu	2575
Qy	1263	TACATGAGCGCTGGGACTTTGAGAGCAGCATTATTGTCCCAGCAAGTTCTCA---	1319
Db	2576	---LysProGlnGluLeuValThAlaGluValAlaAlaProSerThrSerSerSer	2593
Qy	1320	-----GCACCTCCAAACTGAGAAAAATAGGCAAGGCTTCTCTCTGT	1361
Db	2594	AlaThrSerSerProGluGlyProSerProAlaArgProProArgArgThrSerAla	2613
Qy	1362	AT-----CTCCTGGGAAGTAGTGGAGGCGACATGTGCCTGCTGCTGCCAGGCTGA	1414
Db	2614	AspValGluIleArgGlyGlnGlyThrGlyArgProGlyGlnProProGlyPro-----	2631
Qy	1415	GGAAGCAGACGAGCAGCACTGTCCCAGCTTGGCTGAGCAGATGATCCTGGAGGAGTT	1474
Db	2632	-----LysValLeuArgLysLeu-----ProGlyArgLeu	2641
Qy	1475	TGAAGGTGCTGTCGCAGATTCTCCACATCGAGTTCAAGTCCAGGGGCTGAAAATTGA	1534
Db	2642	-----ValThrValValGluGluLysGluLeuValGlnArgArgGlnGlnArgGly	2659
Qy	1535	GTAGCTGCAAGCTGGTAAAGGGG-----AATGCTGTGGCAAGCCCTCAG	1579
Db	2660	AlaAlaSerThrLeuValProGlyValSerGluThrSerAlaSerProGlySerProSer	2679
Qy	1580	-----CCCTGGGAATCTGCA	1594
Db	2680	ValArgSerMetSerGlyProGluSerSerProProlleGlyGlyProCysGluAlaAla	2699
Qy	1595	CCGAGGAGCTGGTCCAGGGAGGAGCAGCCGCCATCATGCGCCAGG-----	1645
Db	2700	ProSerSerLeuProThrProProGlnGlnProPheIleAlaArgArgHisIleGlu	2719
Qy	1645	-----	1645
Db	2720	LeuGlyValThrGlyGlyGlySerProGluAsnGlyAspGlyAlaLeuLeuAlaIleThr	2739
Qy	1646	---TCAGCTGTGAGGTGTGAGTATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT	1702
Db	2740	ProProAlaValLysArgArg-----ArgGlyArgProPro	2751
Qy	1703	AAA-----GTGCAAGCAGGAGCAGCCACCTGGGGATGGACACTGGCCCT	1747
Db	2752	LysLysAsnArgSerProAlaAspAlaGlyArg-----GlyValAspGluAlaPro	2768

```

RESULT 9
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chivria, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)

```

```

US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

```

```

Alignment Scores:
Pred. No.:          0.00853
Score:             118.50
Percent Similarity: 33.2%
Best Local Similarity: 23.61%
Query Match:       2.96%
DB:                4
Length:            3118
Matches:           127
Conservative:      52
Mismatch:          165
Indels:            196
Gaps:              27

```

```

US-09-743-237-1 (1-2241) x US-09-579-181-1 (1-3118)

Qy 577 CCGTGTGCAAGTTCCCATCATCCAGGAGCGAGAGGCGCTCCA-----GCTGGCCCTC 630
Db 2300 ProValIleSerAlaHisGlnThrArgSerThrThrProProArgCysSer-ProAl 2319
Qy 631 GGAAGAAAGACTC---CAGCCCCATGGTGATTGTCAGCTGAAGAGGCGCCAGATGC 687
Db 2319 aArgGluArgValProArgPro-----AlaProArgProArgProThr-- 2333
Qy 688 TCCTGCATAGACAACACTGTGGCGCGAGGAGCTCAAGCGCTCCATCTCTGCT-----TCCTC 741
Db 2334 -----ProAlaSerAlaProAlaAlaIleProAlaLe 2344
Qy 742 AGTACGATGACCAGAGCAGTTTCCC---TCAGTCAGAGCTCCCTTAAGCCCAATGACACACTT 798
Db 2344 uValProValProValSerAlaProValProIleSerAla-ProAsnProIleThrIleL 2364
Qy 799 TAGTGGGAACACTCTGCCAGTACCAGCGAAGTTAAATCTCATCACACAGGTTGTAATAAG 858
Db 2364 euProValHisIleLeuProSerProProProProSerGlnIleProProCysSerSerP 2384
Qy 859 GAGCTCTC-----CCATCAGCTAG 876
Db 2384 roAlaCysThrProProAlaCysThrProProAlaHisThrProProProAlaG 2404
Qy 877 TCAATGGGCTGCTTTTCCCTCTGGAGCTCCTCTGCAAGGCGCACCCAAAATAACTCTGT 936
Db 2404 lnThrCysLeuValThrProSerSerProLeuLeuLeuGlyProProSerValProIleS 2424
Qy 937 CTGGGTACTGTGACTGCTTCTCCAGCGGGGACTTCT----- 972
Db 2424 er-AlaSerValThrAspLeuProLeuGlyLeuArgProGluAlaGluLeuCysAlaGln 2443

```

```
QY 1748 -----CCTGTCCTCGGG 1759
Db 2769 SerSerThrLeuLysGlyLysThrAsnGlyAlaAspProValProGly 2784
RESULT 10
US-09-252-991A-28481
; Sequence 28481, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28481
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28481
Alignment Scores:
Pred. No.: 0.00421 Length: 567
Score: 117.50 Matches: 129
Percent Similarity: 34.43% Conservative: 77
Best Local Similarity: 21.57% Mismatches: 265
Query Match: 2.93% Indels: 127
DB: 4 Gaps: 24
US-09-743-237-1 (1-2241) x US-09-252-991A-28481 (1-567)
QY 527 AGGCGGAGCCCTGAAGCGAGCTTTCAGCGCCCTCGCTCAGGAATCCTGTTGCAA 586
Db 19 ArgArgMetProMetArg-----ValProMetProGlyAlaGluSerMetSerMetArg 36
QY 587 GTTCCCATCATCCAGGAGGAGGAGGCGCTCCAGCTCCCTCGG-----AAGAAGACTC 643
Db 37 ProProMetAlaProMetLeuArg-PheThrThrSerMetProThrProArgProAspSe 56
QY 644 CAGCCCCATGTGATTGTCAGCTGAAGAGGCGGCCAGATGCTCTGC-----692
Db 56 rSerValThrValSerAlaValAluLysProGlyMetAsnThrSerCysGlnThrSerAl 76
QY 693 -----ATAGACAACCTGT---GGCGGAGGAGGAGCTCAAAGCGCTCCATCTGCTCCCTCA 742
Db 76 aSerLeuIleProSerCysThrGlyArgLeuArgSerArgAlaLeuAlaArgMetArgGl 96
QY 743 GTACGATGACAGAGCAGTTTCCCTCAGTCAGAGCTCCCTAAGCCAATCACAACCTTAGT 802
Db 96 yArgAsp-----ArgProArgProSerSerGluThrSe 107
QY 803 GGAAGACTTCTGCCAGTACACAGCAAGTTAAATCTCATCACACAGGTGATAATGGAGC 862
Db 107 rThrThrThrLeuPro---ProTrpTrpLeuAlaAlaSerPheThrValProLeuAlaGl 126
QY 863 TCTCCCATCAGCTGTCAAT-----GGGCTGCCCTTCCCTCGAGCGCTGCTCAAGG 916
Db 126 yLeuProAlaAlaArgArgSerAlaGlyValSerMetProTrpSerMetLeuLeuArg-- 145
QY 917 GCCACCAAAATAACTCTGCTGGG-----TACTGTGACTCTCTCCAGCGGGA 967
Db 146 -----ThrMetTrpValSerGlySerLeuSerPheSerIleThrLeuLeuSerSerSe 163
QY 968 CTTCTGCAACAGCTGCAACAACTGCGCCCATGAGTCGAGCGCTTCAAAGCCAT 1027
Db 163 rValCysSerProCysArgCysSerSer-----172
QY 1028 AAAGCGGTCTCTGTAGAAATCTTGAAGCTTTCCCAACCAAAATGGGAAA-----1079
```

```
Db 173 ----ThrCysLeuProArgArgValAlaValSerArgThrSerLeuGlyLysArgLeuLy 191
QY 1080 ---GGCCGCTCTGGGAGCTGCTAAACTTCGACACAGCAAGAGGTGCAAC-----1124
Db 191 sMetLysLeuIleGlySerMetArgIleAspIleThrAspSerCysArgSerArgVal 211
QY 1125 -----TGTAAGCGCTCAGGCTCCCTCGCTCAGCAAGCACTACTGTGA 1159
Db 211 uCysSerSerTrpAlaArgAlaCysCysArgArgSerTrpArgLeuGlySerSerCysSe 231
QY 1160 G-----TGCTATGAGGCAAAATCATGTGTTCTTCCATTGCCAATGATGCTGTG 1210
Db 231 rProSerTrpProSerMetAlaTrpValIleThrSerSerProThrArgLeuIleArgLe 251
QY 1211 CAAAACATATGAA-----CAAAAGTCAGAACGAAATGCTGAT 1249
Db 251 uSerIlePheSerThrLeuThrArgIleGluProAspSerProSerProArgThrGlyAl 271
QY 1250 GAGCACACCCC-----ACTACATGAGCGCTGGGACTTTGAGAGCAGCCATTA 1297
Db 271 aThrAla-ProLeuAlaValAlaAlaThrAlaCysAlaGlyAlaLeuGlyAlaGlyL 291
QY 1298 TTTGTCGCCCGCAAGTTCTCAGGACCTCCCAAACTGAGAAAAATAGCAGCCCT----1353
Db 291 euProGlyAlaProSerAsnArgProLysAlaArgSerSerAlaGlyGlyProGlyA 311
QY 1354 -----T 1354
Db 311 laSerSerLysArgProAsnAlaSerLeuAlaAlaGlyAlaProGlyAlaSerSerLysS 331
QY 1355 CTCCTGTATCTCCTGGGAAGTAGTAGGAGGCCACATGTGCTGCTCCCTCGCTGGCCAGGGTGA 1414
Db 331 erProAsnAlaProGlyAlaAlaAlaArgSerGlyAlaProGlyProSerSerLysLysP 351
QY 1415 GGAACGAGAGGAGGAGCTGTTCCCAAGCTTGGCTGAGC---AGATGATCTCGGAGGA 1471
Db 351 roLysAlaSerSerAlaValSerLeuAlaGlyLeuAlaGlyAlaSerSerAsnArgP 371
QY 1472 GTTTGGAAGGTGCTGTCGCGAGATTCTCACATTCAGTTCAAAGTCCAAAGGGGCTGAAAT 1531
Db 371 roLysProSerSerAlaAlaSerCysSerGlyThrArgSerGlnProArg-----ArgA 389
QY 1532 TAGTAGGCTGCAAGCTGTGTAAGGGGAATGCTGTGGCAAGCCCTCAGCCCTGGGAATCT 1591
Db 389 rgSerThrValSerSerAsnGlyThrPro-GlySerGlyArgProLeuArgTrpAlaSer 408
QY 1592 GCACCGAGGAAGCTGGTCCCGAGGAGGAGCAGAGCGCGCGCATCATGCGCC---AGGTCA 1648
Db 409 SerArgArgCysAsnThrSerArgAlaSerArgLysGluLeuIleMetSerGlyArgArg 428
QY 1649 GCTGTGAGGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1708
Db 429 CysIle-----TrpProGlnArgMetArgSerSerArgVal 440
QY 1709 CAAGCAGCAGCAGCAGCCCTGGGATGGACACTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1766
Db 441 SerSerSerTrpValThrAlaValMetSerSerLysProAsnIleProLeuGlu-ProLe 460
QY 1767 -----TCTGGGGACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1816
Db 460 uIleGluCysAlaAlaArgLysIleProCys-----470
QY 1817 TTATGTGTTTGTCTTCAAAATCTAGTAGTACCTCCATTCAGTTATTATGAGCCAGCC 1876
Db 471 -----SerValSerLeuSerGlyArgSerArgSerArgProIleArgArgSerSe 487
QY 1877 TCAAGTTAGAGAGCTAGCTCTTCTTTCAGGTGGAGCTCTGCCCAATACATACATACAGTCAAG 1936
Db 487 rArgValAlaArgCysSerSerValSerSerLysAsnThrTrpArgAsnTrpLeuIleSe 507
QY 1937 GTGGCCATCAGGGGTTTTTCCAGGCCAGGCTGTGACAGGAGATATGAGGGGGGTGCG 1996
Db 1937 GTGGCCATCAGGGGTTTTTCCAGGCCAGGCTGTGACAGGAGATATGAGGGGGGTGCG 1996
```


Db 507 rTrp-----GlyMetGlyProProArgGlyAsnGlnGlyArgThrLeuAr 525
QY 1997 GTTAGAGCTGGTTGTTT---GGATTTTTTGGCTTTTCTTCCTGTTATTTCTGCTTG 2053
Db 525 gValThrSerSerSerPheGlyGlySerAsnGlyPheThrSerGlnProValAlaProAl 545
QY 2054 AAGTGAGAAACTGTCTCTGTCCTCAACCTTTTCTGCATAATTAAT 2099
Db 545 aLeuArgProCysCysPheIleAlaSerLeuAspSerValValSer 560

RESULT 11

US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Yaciuk, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Alignment Scores:
Pred. No.: 0.0105 Length: 2972
Score: 117.50 Matches: 127
Percent Similarity: 33.27% Conservative: 52
Best Local Similarity: 23.61% Mismatches: 165
Query Match: 2.93% Indels: 196
DB: 4 Gaps: 27

US-09-743-237-1 (1-2241) x US-09-579-181-2 (1-2972)

QY 577 CTTGTTCCAGTTCCCATCATCCAGAGGAGGAGGCTCCA-----GCTGCGCTC 630
Db 2154 ProValIleSerAlaHisGlnThrArgSerThrThrThrProArgCysSer-ProAl 2173
QY 631 GGAAGAAAGACTC---CAGCCCATGTGTGTTCTCAGCTGAAGAGGAGGCCAGATGC 687
Db 2173 aArgGluArgValProArgPro-----AlaProArgProArgProThr-- 2187
QY 688 TCTGCATAGAACTGTGGCGGAGGAGCTCAAGCGCTCCATCTGCT-----TCCTC 741
Db 2188 -----ProAlaSerAlaProAlaAlaIleProAlaLe 2198
QY 742 AGTACGATGACCAGACAGTTTCC-----TCAGTCAGAGCTCCCTAACCAATGACACTT 798
Db 2198 uValProValProValSerAlaProValProIleSerAla-ProAsnProIleThrIleL 2218
QY 799 TAGTGGGAAGACTTCTCCAGTACCGAAGTTAAATCTCATCACACAGTTTGATAATG 858
Db 2218 euProValHisIleLeuProSerProProProProSerGlnIleProProCysSerSerP 2238
QY 859 GAGCTCTC-----CCATCAGCTG 876
Db 2238 roAlaCysThrProProAlaCysThrProProAlaHisThrProProProAlaG 2258
QY 877 TCAATGGGGTGCCTTTCCTCTGGAGCTGCTGTGGAAGGGCCACCCAAATAACTCTGT 936
Db 2258 lnThrCysLeuValThrProSerSerProLeuLeuGlyProProSerValProIles 2278
QY 937 CTGGGTACTGTGACTGCTTCTCCAGCGGGACTTCT----- 972
Db 2278 er-AlaSerValThrAsnLeuProLeuGlyLeuArgProGluAlaGluLeuCysAlaGln 2297

QY 973 -----GCAACAGCTGCAGCTGCAACAACCTGCGCC 1002
Db 2298 AlaLeuAlaSerProGluSerLeuLeuAlaSerValAlaSerSerGluThrSerSer 2317
QY 1003 ATGAGCTCGAGCGCTTCAAAGCCATAAAGCGTCTCTGTATAGAAATCTCTGAAGCTTCC 1062
Db 2318 LeuSerLeuValProProLysAspLeuLeuProValAlaValGluIleLeuProValSer 2337
QY 1063 AACCAAAATGGGAAAGCGCTGGGAGCTGCTAAACTTCGACACAGCAAAAGGTGCA 1122
Db 2338 GluLys-----AsnLeuSerLeuThrProSerAlaProSerSerLeu 2350
QY 1123 ACTGTAAGCGCT-----CAGGCTGCGCTGA 1146
Db 2351 ThrLeuGluAlaGlySerIleProAsnGlyGlnGluAlaProAspSerAlaGlu 2370
QY 1147 AGAATCTACTGTGAGTGTCTATGAGGCCAAATCA-----TGTGTT----- 1185
Db 2371 GlyThrThrLeuThrValLeu---ProGluGlyGluLeuProLeuCysValSerGlu 2389
QY 1186 -----CTTCATTTGCAATGCATT----- 1205
Db 2390 SerAsnGlyLeuGluLeuProProSerAlaAlaSerAspGluProLeuGlnGluProLeu 2409
QY 1206 ---GCTTGCATAAACTATGAAGAAAGTCCAGAACGAAATAATGCTGATGACACACCCAC 1262
Db 2410 GluAlaAspArgThrSerGluGluLeuThrGluAlaLysThrProThrSerSerProGlu 2429
QY 1263 TACATGAGCGCTGGGACTTTGAGAGCAGCCATTTTGTCCGCCAGCAGCTTCTCA--- 1319
Db 2430 ---LysProGlnGluLeuValThrAlaGluValAlaAlaProSerThrSerSerSer 2447
QY 1320 -----GGACCTCCAAACTGAGAAAATAGCAGGCGCTTCTCTGT 1361
Db 2448 AlaThrSerSerProGluGlyProSerProAlaArgProProArgArgThrSerAla 2467
QY 1362 AT-----CTCCTGGGAAGTAGTGGAGGCCACATGTGCTGCTGCGCCAGGCTGA 1414
Db 2468 AspValGluIleArgGlyGlnGlyThrGlyArgProGlyGlnProProGlyPro----- 2485
QY 1415 GGAAGCAGACGAGGACACTGTTCCTCCAGCTGAGCAGATGATCTCTGGAGGAGTT 1474
Db 2486 -----LysValLeuArgLysLeu-----ProGlyArgLeu 2495
QY 1475 TGAAGGTGCTGTGCAGACTTCCACATCGAGTTCAGTCCAAAGGGCTGAAAATGA 1534
Db 2496 -----ValThrValValGluGluLysGluLeuValGlnArgArgGlnArgGly 2513
QY 1535 GTAGCGTCAAGCTGTGTAAGGGG-----AATGCTGTGGCAAGCCTCAG 1579
Db 2514 AlaAlaSerThrLeuValProGlyValSerGluThrSerAlaSerProGlySerProSer 2533
QY 1580 -----CCTGGGATCTGCA 1594
Db 2534 ValArgSerMetSerGlyProGluSerSerProProIleGlyProCysGluAlaAla 2553
QY 1595 CCAGGAGGCTGTGGTCCGAGGAGGAGGAGGCGCATCATGCGCAGG----- 1645
Db 2554 ProSerSerLeuProThrProProGlnGlnProPheIleAlaArgArgHisIleGlu 2573
QY 1645 ----- 1645
Db 2574 LeuGlyValThrGlyGlySerProGluAsnGlyAspGlyAlaLeuLeuAlaIleThr 2593
QY 1646 ---TCAGCTGTGAGTCTGAGTGATCTGCATGGTACTGGCCAGCGCTACTCAAGTATCCT 1702
Db 2594 ProProAlaValLysArgArg-----ArgGlyArgProPro 2605
QY 1703 AAA-----GTCAAGCAGGAGGAGCCACCCTGGGGATGGACACTGGCCCT 1747
Db 2606 LysLysAsnArgSerProAlaAspAlaGlyArg-----GlyValAspGluAlaPro 2622
QY 1748 -----CCTGTCCCTGGG 1759

Qy 1550 CCAGCTTGACGCTACTCAATTTTCAGCCCTTGGACTGGAATCGATGCGAATCTG 1491
Db 180 GlyAlaCysThr-SerSerAlaAlaGluPro---ProProSerSerThrArgArgGlySer 198
Qy 1490 CGACAGGACCTTCCAACTCCCTCCAGGATCATCTGCTAGCCCAAGCTTGGGACAGTG 1431
Db 198 rSerGlnSerLeuProGlyAlaArgArgIleValGlnSerAla-----Cys 213
Qy 1430 CTCTGCTGCTCTCTCCACCC-----TGGCCAGGAGGAGGACGACATGCGC 1383
Db 213 sArgCysArgAlaCysGlyAlaSerGlyArgArgTrpAlaSerGlyAsnGlnArgLeuSe 233
Qy 1382 CTCACCTACTTCCAGGAGATACAGGAGAGGCTCGCTATTTTCTCAGTTTGGAGG 1323
Db 233 rGlnThrArgPro-ArgValSerGlyArgPro----- 244
Qy 1322 TCCTGAGAACTTGGCTGGGACAAATAATGGCTGCTCTCAAAAGTCC---CAGGCTCCAT 1266
Db 245 -----GlyAlaSerArgArgProCysGlnArgProS 255
Qy 1265 GTAGTGGGTGCTCATCAGCATTTTTCGTTCTGGACTTCTTCATAGTTTTCGAAGC 1206
Db 255 erPro-----SerAspGlyAlaSerTrpLeuSerSerValSerSerAlaCysArgS 272
Qy 1205 AATGCATTTGCAATGGAAGAACACATGATTTTGGCCCTCATAGCACTCACAGTAGTTCTT 1146
Db 272 erSerCysAlaArgValArgArg-----GlnProValGlySerSerLyss 288
Qy 1145 CAGCAGCCTGAGGCTTACAGTTGCAACCTTTGCTGT-----GTGGAAG 1101
Db 288 erTrpAlaSerSerTrpAlaCysCysAlaSerCysArgArgSerCysAlaAlaSerG 308
Qy 1100 TTTAGCAGCTCCAGAGCGC-----CTTTCCCATTTTGGTTGGAAGCTTCAGG 1050
Db 308 lYasnSerAlaProAspGlyArgProMetSerSerPro----- 320
Qy 1049 ATTTCTATCAAGACACGCCCTTTATGGCTTTGAAGCGCTCGAGCTCATGGCGAGTTGTT 990
Db 320 ----- 320
Qy 989 GCAGCTCAGCTGTGTCAGAGTCCC-----CGCTGGAGAGAGCAGTC 948
Db 321 -----ProSerCysSerGlnArgProAlaGlyGlnGlyArgSerArgTrpLysergSerA 339
Qy 947 ACAGTACCCAGACAGAGTTATTTTGGGTGGCCCTTCAGAGCAGGTCAGAGGGAAGGC 888
Db 339 rg-----TrpIleCysSerGlySerGlnAlaArgValSerArgA 352
Qy 887 AGCCCCATTGACAGCTGATGGGAGAGCTCCATTATCAACCTGTGTGATGATTTAACTT 828
Db 352 rg-----AspValGlnS 356
Qy 827 CGCTGGTACTGGCAGAGTCTTCCACATAAGTTGTCTATGGCTTAGGAGCTCTGACTG 768
Db 356 erSerAlaAlaAlaSerSerSerGluLeu-----AlaAlaAlaGlyCysGlnAlas 373
Qy 767 AGGGAATCTGCTCT-----GGTCATCGTACTGAGGAGCAGATGGAGCGCTTT 720
Db 373 erAlaThrCysSerThrArgArgProGlyArgArgGlyArgGlnProAsp----- 389
Qy 719 GAGCTCCCTCGGCCACAGTTGTCTATGCAGAGCATCTGGCGCTCTCTTCAGCTGACA 660
Db 390 --SerProGlyGlnArgSerSerAlaCysSerThrSerArgArgLeu---AlaProS 408
Qy 659 AATCACCATGGGCTGGAGTCTTTCTTCAGGAGGACGCTGGAGGCTCTCTGCTGCT 604
Db 408 erAlaPro-----ArgArgSerSerAlaSerArgSerProMetPro 422

RESULT 14

US-09-252-991A-17335

; Sequence 17335, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17335
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17335

Alignment Scores:

Pred. No.: 0.00675 Length: 456
Score: 115.00 Matches: 67
Percent Similarity: 29.53% Conservative: 21
Best Local Similarity: 22.48% Mismatches: 74
Query Match: 2.84% Indels: 136
DB: 4 Gaps: 14

US-09-743-237-1 (1-2241) x US-09-252-991A-17335 (1-456)

Qy 1782 GCAGGAGTCCCGAGAGGGCTCCCGCAGGACAGAGGCGCAGTGTCCATCCCGAGGGTG 1733
Db 277 AlaGlyAlaGlyArgAlaAlaGlyProGlyThrGlyGlnProArgArgGlyHisArgLeu 296
Qy 1722 GCTGCTGCTGCTGCTGCTTTAGGATACCTTCAGTAGGTGCGCCAGTACCATGCAGATCAC 1663
Db 297 AlaThrPro----- 299
Qy 1662 TCAGACCTCACAGCTGACCTGGCCATGATGCGCGCTCTGCTCTCTCCCTGGCAGCAGC 1603
Db 300 -----GlyArgProAlaGlyProGlyArgSer 308
Qy 1602 TTCTCTCGTGCAGATTCCCGAGGCTGAGGCTTGGCCAGGACATTCCTTTTACAGCTTG 1543
Db 309 LeuProAlaAlaGlyGlyGlyGlyGlyGlyLeuHisArg----- 321
Qy 1542 CACGCTACTCAATTTTCAGCCCTTGGACTTGAACCTCGATGTGAGAACTCTCGCAGAGC 1483
Db 322 -----GlnCysGlyAspProAlaLeuGly 329
Qy 1482 ACCTTCCAACTCCTCCAGGATCATCTGCTCAGCCAAAGCTTGGGAAACAGTGTCTCTGCT 1423
Db 330 GlyGlyGlyGlySerPro-----AlaArgArgProAlaLeuGlyHisGlySerArgVal 347
Qy 1422 CTGCTTCTCCTCACCCTGGG-----CCAGCAGGACAGG----- 1393
Db 348 Val-----ArgProGlyCysProGlnHisAlaLeuArgProAlaArgArgAlaAlaAsp 365
Qy 1392 -----CACATGTGCTCCACTTCTCCAGGAGATCAGCAGGAAGG 1351
Db 366 AspArgGlyAspArgThrGlnLeuGlyProThrSerMetPro-----ValArgGlu 382
Qy 1350 CTTGCTCTATTTTCTCAGTTTGGAGTCTCTAGAACTTGGCTGGGACAAATTAATGGC 1291
Db 383 ProAlaProArgArgSer-----SerThrAlaSerGly 393
Qy 1290 TGCTCTCAAGTCCCGAGGCTCCATGTAGTGGGTGTGCTCATCAGCATTTTTCGTTCTG 1231
Db 394 Cys-----SerProSerSerSerCysSerAlaAlaCysCysGlyArgPhe----- 408
Qy 1230 GACTTCTTCATAGTTTTCACCAAGCAATGATTTTGGCAATGGAAGAACACATGATTTGG 1171
Db 409 -----ArgTrpArg-----AlaTrp 413
Qy 1170 CCTCATAGC---ACTCACAGTAGTTTCTTCAGGACGCTGAGCGCTTACAGTTGCACCTT 1114

```

Db      414  ProGlyCysTrpSerIleProGlySerSerTrpSerIleSerArgProSerAlaThrSer 433
QY      1113  TGCCTGTGCGAAGTTTACAGCTCCAGAGCGGCTTTCCCATTTTGGTTGGAAGCTT 1054
Db      434  LeuAlaThr-----
QY      1053  CAGGATTTCTATCATCAGACACAGCGCTTTATGGCTTTGAAGCGCTCAGAGCTCATGGCGCAGGT 994
Db      437  -----AlaGly 438
QY      993  TGTTCAGCTGCAGCTGTGCGAAGTCCCGCTGGAGAGCAGTCCACAGTACC 940
Db      439  ---CysArgCysTrpCysThrArgProProProTrpArgSerSerProThr 455

RESULT 15
US-09-252-991A-23774
; Sequence 23774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23774
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23774

Alignment Scores:
Pred. No.:      0.0131      Length:      798
Score:          113.50      Matches:     115
Percent Similarity: 33.48%      Conservative: 42
Best Local Similarity: 24.52%      Mismatches:  136
Query Match:      2.83%      Indels:      176
DB:               4         Gaps:         27

US-09-743-237-1 (1-2241) x US-09-252-991A-23774 (1-798)
QY      554  CCAGGCCCTCTGGCTCAGGAATCCTGTTGCAAGTTCCCATCCAGGAGGCAGGA 613
Db      91   ProAlaAlaSerGlyProGly-AlaCysCysSerSerProAla-ProGlySerThrArg 110
QY      614  GGCTCCAGCTGCCCTCGGAAGAAGACATCCAGCCCATGGTGTGTCAGCTGAAGG 673
Db      110  rgProMetaAlaAla-GlySerThrArg-----SerArgArg 121
QY      674  AGGCGCCAGAGCTCTGCTAGACACAACTGTGGCGGAGGAGCTCAAGCGCTCATCT 733
Db      122  ArgPheProArgSerIle-----ProAspGlySerCysSerThrProSer 136
QY      734  GCTTCCTCAGTACGATGACCAGCAGCATTTCCCTCAGTCAGAGCTCCCTAAGCCATGAC 793
Db      137  ArgSerArgTrpThrAlaGlyTyrArgLeuProAlaCysAlaTrpPro-----AlaSer 154
QY      794  AACTTTAGTGAAGACTTCTGCGAGTACCGAGGAGTTAAATCTCATCACACAGGTTGA 853
Db      155  ThrValArgGlyGlySerGlyCysSerThrArgCysValArgGly-----ThrGly--- 171
QY      854  TAATGAGCTCTCCCATCAGCTGT-----CAATGGGCG 886
Db      172  ---TrpSerAlaSerArgProCysMetProThrGlySerThrProThrMetArgProGly 190
QY      887  TGCCTTTCCCTGGACCTGCTCTGCAAGGGCCACCACCAAAATAACTCTCTCTGGTACTG 946
Db      111  : : : : :

```

Search completed: July 24, 2003, 14:38:11
Job time : 112.575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:37:43 ; Search time 112.946 Seconds
(without alignments)
6298.674 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 4008
Sequence: 1 taccctgtggtggcccg.....aaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09743237/runat.24072003_113103_17652/app_query.fasta_1.4750
-DB=A_Geneseq_19Jun03 -Qfmt=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09743237.ecgn.1.1.376.runat.24072003_113103_17652 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1599	39.9	295	21	AAV68463
2	1251	31.2	299	21	Mouse testis speci
3	639	15.9	528	24	Human testis speci
4	503	12.5	147	22	Human CGDD protein
5	503	12.5	147	22	Human testicular a
6	440.5	11.0	950	22	Human reproductive
7	405.5	10.1	438	20	Drosophila melanog
8	362	9.0	280	22	Caenorhabditis ele
9	346	8.6	251	22	Novel human diagno
10	248	6.2	53	22	Human polypeptide,
11	226.5	5.7	609	24	Human colon cancer
12	186	4.6	243	22	Histone deacetylase
13	140.5	3.5	4561	22	Drosophila melanog
14	140.5	3.5	9222	22	Novel human diagno
15	136.5	3.4	966	23	Novel human diagno
16	131.5	3.2	1061	20	Human TSPI domain
17	131.5	3.3	1192	24	Human N-methyl-D-a
18	131.5	3.3	1207	24	Human kiellin-like
19	131.5	3.3	1477	24	Human kiellin-like
20	131.5	3.3	1535	24	Human kiellin-like
21	131.5	3.3	1593	24	Human kiellin-like
22	130.5	3.2	19938	24	Streptomyces virid
23	127	3.1	961	23	Human TSPI domain
24	127	3.2	1774	23	Human REMAP-6 prot
25	126.5	3.2	1454	22	Human polypeptide
26	126	3.1	19938	24	Streptomyces virid
27	125.5	3.1	1057	24	Human kiellin-like
28	125	3.1	961	23	Human TSPI domain
29	124.5	3.1	1081	21	Human N-methyl-D-a
30	124.5	3.1	1081	23	Human NMDA recepto
31	124.5	3.1	1081	24	Human N-methyl-D-a
32	124	3.1	1251	24	Human kiellin-like
33	124	3.1	1342	24	Human kiellin-like
34	124	3.1	1512	24	Human kiellin-like
35	124	3.1	1570	24	Human kiellin-like
36	124	3.1	1628	24	Human kiellin-like
37	124	3.1	3680	22	Human kiellin-like
38	123	3.1	19938	24	Drosophila melanog
39	122.5	3.1	1017	22	Streptomyces virid
40	122.5	3.1	1615	22	TutD protein #4.
41	122.5	3.1	19938	24	Protein #3 encoded
42	121	3.0	919	24	Streptomyces virid
43	120.5	3.0	966	23	Human extracellular
44	120.5	3.0	1212	20	Human TSPI domain
45	120.5	3.0	1235	23	Human N-methyl-D-a
					Human tumour suppr

ALIGNMENTS

RESULT 1
AAV68463
ID AAV68463 standard; Protein; 295 AA.
XX AAV68463;
AC AAV68463;
XX
XX 25-APR-2000 (first entry)
DT
XX
DE Mouse-testis specific factor tesmin SEQ ID NO:4.
XX
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
XX Mus musculus.
XX
XX
PN WO200004147-A1.
XX

PD 27-JAN-2000.
 XX PF 16-JUL-1999; 99WO-JP03859.
 XX PR 17-JUL-1998; 98JP-0219856.
 XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX WPI: 2000-147785/13.
 DR N-PSDB; AAZ88155, AAZ88157.
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 XX Claim 1; Page 47-49; 63pp; Japanese.
 XX The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX SQ Sequence 295 AA;

Alignment Scores:
 Pred. No.: 2,96e-142 Length: 295
 Score: 1599.00 Matches: 295
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 39.90% Indels: 0
 DB: 21 Gaps: 0

US-09-743-237-1 (1-2241) x AAY68463 (1-295)

QY 651 ATGGTGATTGTGACGTGAAGAGGCGGCCAGATGCTCTGCATAGACAACHTGTGGCGG 710
 Db 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
 QY 711 AGGGAGCTCAAGCGCTCCATCTGCTTCTCAGTACGATGACGAGAGTTTCCCTCAG 770
 Db 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAsnGlnSerPheProGln 40
 QY 771 TCAGAGCTCCCTTAAGCCAATGACAATTTAGTGGGAAGACTTCTGCCAGTACCAGCGAAG 830
 Db 41 SerGluLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
 QY 831 TTAATCTCATCACAGATTGATATGAGAGCTCTCCCATCAGCTGTCATGGGCTGCC 890
 Db 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
 QY 891 TTTCCCTCTGGACCTGCTGCAAGGCGCCACCAAAATAACTCTGCTGGGTACTGTGAC 950
 Db 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
 QY 951 TCTCTCTCAGCGGGAGCTTCTGCAACAGCTGACGCTGCAACAACCTGCGCCATGAGTCT 1010
 Db 101 CysPheSerSerGlyAspPheCysAsnSerCysSerCysAsnAsnLeuArgHisGluLeu 120
 QY 1011 GAGCGCTTCAAGCCATTAAGCGGTGCTTGATAGAAATCTCTGAAGCTTTCCACCAAAA 1070
 Db 121 GluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheGlnProLys 140
 QY 1071 ATGGGAAGGCGCTCTGGAGCTGCTAAACTTCACACAGCAAGGGTGCAACTGAAG 1130
 Db 141 MetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGlyCysAsnCysLys 160
 QY 1131 CGCTCAGGCTGCTGGAAGCACTACTGCTAGTCTATGAGGCCCAAAATCATGTGTTCTTCC 1190
 Db 161 ArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIleMetCysSerSer 180
 QY 1191 ATTTGCAATGCTGCTTGGCAAAACTATGAAGAAGTCCAGAACGAAATAATGCTGATG 1250

Db 181 IleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArgLysMetLeuMet 200
 QY 1251 AGCACACCCACTACATGAGCTGGGAGCTTTGAGAGCAGCATTTATTTGTTCCCGAGCC 1310
 Db 201 SerThrProHisTyrMetGluProGlyAspPheGluSerHisTyrLeuSerProAla 220
 QY 1311 AAGTTTCTCAGGACCTCCAAACTGAGAAAAATAGCAGCGCTTCTCTGTATCTCTCGG 1370
 Db 221 LysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSerCysIleSerTrp 240
 QY 1371 GAAGTAGTGGAGGCCACATGTCCTGCTGCTGGCCCGAGGTGAGAAAGCAGAGGAGG 1430
 Db 241 GluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluAlaGluGlnGlu 260
 QY 1431 CACTCTTCTCCCAAGCTTGGCTGAGCAGATGATCTCTGGAGAGCTTTGGGAAGGTGCTGTCG 1490
 Db 261 HisCysSerProSerLeuAlaGlnMetIleLeuGluGluPheGlyArgCysLeuSer 280
 QY 1491 CAGATTCTCCACATCGAGTTCAAGTCCCAAGGGGCTGAAAATTGAG 1535
 Db 281 GlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295

RESULT 2
 AAY68464
 ID AAY68464 standard; Protein; 299 AA.
 XX AC AAY68464;
 XX DT 25-APR-2000 (first entry)
 XX DE Human testis specific factor tesmin SEQ ID NO:5.
 XX KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 XX KW germ cell differentiation; sterility.
 XX OS Homo sapiens.
 XX PN WO200004147-A1.
 XX PD 27-JAN-2000.
 XX PF 16-JUL-1999; 99WO-JP03859.
 XX PR 17-JUL-1998; 98JP-0219856.
 XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX WPI: 2000-147785/13.
 DR N-PSDB; AAZ88156.
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 XX Claim 1; Page 50-52; 63pp; Japanese.
 XX The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX SQ Sequence 299 AA;

Alignment Scores:
 Pred. No.: 2,69e-109 Length: 299
 Score: 1251.00 Matches: 228
 Percent Similarity: 84.95% Conservative: 26
 Best Local Similarity: 76.25% Mismatches: 41
 Query Match: 31.21% Indels: 4


```
DB:                21                Gaps:                1
US-09-743-237-1 (1-2241) x AAY68464 (1-299)
QY 651 ATGGTATTTGTCAGCTGAAGAGCGGCCAGCATGCTCTGCATACACAACTGTGCGCGG 710
DB 1 MetValIleCysGlnLeuLysGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
QY 711 AGGAGCTCAAGCGCTCCATCTGCTCTCAGTACGATGATGACAGAGCAGTTTCCCTCAG 770
DB 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGln 40
QY 771 TCAGAGCTCCCTAAGCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGCGAAG 830
DB 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
QY 831 TTAATCTCATCACACAGGTGATATGAGCTCTCCATCAGCTGTCATATGGGGTGGC 890
DB 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
QY 891 TTTCCCTCTGGACCTGCTCTGCAAGGCGCCACCCAAATAACTCTGCTGGGTACTGTGAC 950
DB 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGCGGGGACTTCTGCAACAGCTGCAGC-----TGCACAACTGTG 998
DB 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnAsnCysAsnAsnLeu 120
QY 999 CGGCATGAGCTCGAGCGCTTCAAGCCATAAGCGCTGTCTGTATAGAAATCTGAAGCT 1058
DB 121 HisHisAspIleGluArgPheLysAlaIleLysAlaCysLeuGlyArgAsnProGluAla 140
QY 1059 TTCCCAACCAAAATGGGAAGCGCTCTGGGAGCTGCTAACTTCGACACAGCAAGGG 1118
DB 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
QY 1119 TGCAGCTGAAGCGCTCAGCTGCCTGAAGAACTACTGTGAGTGTATGAGCGCAAAATC 1178
DB 161 CysAsnCysArgArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGlnIle 180
QY 1179 ATGTGTCTTCCATTTGCCAAATTCATTTGCTTGCAGAACTATGCAAAAGTCCAGACGA 1238
DB 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200
QY 1239 AAAATCTGATGAGCACACCCACTACATGAGCGCTGGGACCTTTCAGAGCAGCCATTAT 1298
DB 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
QY 1299 TTGTCGCCAGCCAAAGTTCTCAGGACCTCCAAACTGAGAAAATAGGAGCGCTTCTCC 1358
DB 221 LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgArgProSerSer 240
QY 1359 TGTATCTCTGGGAAGTAGTGGAGGCCACATGTGCTGCTGCTGCCCGAGGTGAGGAA 1418
DB 241 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 260
QY 1419 GCAGAGCAGGAGCAGCTGTTCCCAAGCTGGCTGAGCAGATGATCTGAGGAGGTTTGGGA 1478
DB 261 AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluPheGly 280
QY 1479 AGGTGCTGTGCGAGATTTCTCCACATCGAGTTCACATGCCAAGGGGCTGAAATTTAG 1535
DB 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 3
ABG74678
ID -ABG74678 standard; Protein: 528 AA.
XX
AC ABG74678;
XX
DT 10-MAY-2003 (first entry)
XX
DE Human CGDD protein 2369279CD1 SEQ ID 4.
```

```
XX
KW Human: cell growth, differentiation and death; CGDD; cardiant; cancer;
KW cytostatic; neuroprotective; gene therapy; diagnosis; cardiovascular;
KW neurological disorder.
XX
OS Homo sapiens.
XX
PN WO2003014322-A2.
XX
PD 20-FEB-2003.
XX
PF 08-AUG-2002; 2002WO-US25465.
XX
PR 08-AUG-2001; 2001US-311017P.
PR 17-AUG-2001; 2001US-313070P.
PR 17-AUG-2001; 2001US-313071P.
PR 24-AUG-2001; 2001US-314678P.
PR 31-AUG-2001; 2001US-316692P.
PR 07-SEP-2001; 2001US-317913P.
PR 14-SEP-2001; 2001US-322182P.
PR 07-DEC-2001; 2001US-340747P.
PR 20-DEC-2001; 2001US-342761P.
PR 29-MAR-2002; 2002US-369129P.
XX
(INCY-) INCYTE GENOMICS INC.
XX
PI Azimzai Y, Barroso L, Baughn MR, Becha SD, Borowsky ML, Duggan BM;
PI Elliott VS, Emerling BM, Forsythe IJ, Gietzen KJ, Gorvad AE;
PI Graul FC, Griffin JA, Gururajan R, Hatalia AJA, Ison CH, Kable AE;
PI Khan RA, Lee S, Lee SY, Li JX, Reddy R, Richardson TW;
PI Sprague WW, Swarnakar A, Tang YT, Warren BA, Xu Y, Yao MG, Yue H;
PI Yue H;
XX
WPI: 2003-268197/26.
DR N-PSDB; ABQ77417.
XX
PT New polypeptide associated with cell growth, differentiation and death,
PT useful for preparing a composition for diagnosing or treating a disease
PT e.g., cardiovascular or neurological disorders or cancer.
XX
PS Claim 59; Page 201-202; 267pp; English.
XX
CC This invention describes a novel polypeptide associated with cell growth,
CC differentiation and death (CGDD) which has cardiant, cytostatic and
CC neuroprotective activity. The polypeptides of the invention are useful
CC for preparing a composition for diagnosing, or treating a disease or
CC condition associated with decreased expression or overexpression of
CC functional CGDD e.g., cardiovascular or neurological disorders or cancer.
CC The polynucleotides of the invention can be used for gene therapy.
CC ABG74675-ABG74701 represent the human CGDD polypeptides encoded by
CC ABQ77414-ABQ77440.
XX
SQ Sequence 528 AA;
XX
Alignment Scores:
Pred. No.: 3,18e-51 Length: 528
Score: 639.00 Matches: 140
Percent Similarity: 55.76% Conservative: 39
Best Local Similarity: 43.61% Mismatches: 68
Query Match: 15.94% Indels: 74
DB: 24 Gaps: 9
US-09-743-237-1 (1-2241) x ABG74678 (1-528)
QY 756 AGCAGTTTCCCTCAGTCAGAGTCCCTTAAGCCAAATGACAACTTTA----- 800
DB 208 ThrSerGlnProGlnGlnArgLeuIleMetProAlaThrProLeuProGlnIleGlnPro 227
QY 801 -----GTGGGAAGACTTCTG---CCAGTACCAGGAGTTAAATCTC 839
DB 228 AsnLeuThrAsnLeuProProGlyThrValLeuAlaProLalaProGlyThrGlyAsnVal 247
QY 840 -----ATCACAGAGTTTGATAATGAGAGCTCTCCCATCA 872
```

1

Gaps:

21

DB:

US-09-743-237-1 (1-2241) x AAY68464 (1-299)

QY 651 ATGGTATTTGTCAGCTGAAGAGCGGCCAGCATGCTCTGCATACACAACTGTGCGCGG 710

DB 1 MetValIleCysGlnLeuLysGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20

QY 711 AGGAGCTCAAGCGCTCCATCTGCTCTCAGTACGATGATGACAGAGCAGTTTCCCTCAG 770

DB 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGln 40

QY 771 TCAGAGCTCCCTAAGCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGCGAAG 830

DB 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60

QY 831 TTAATCTCATCACACAGGTGATATGAGCTCTCCATCAGCTGTCATATGGGGTGGC 890

DB 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80

QY 891 TTTCCCTCTGGACCTGCTCTGCAAGGCGCCACCCAAATAACTCTGCTGGGTACTGTGAC 950

DB 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100

QY 951 TGCTTCTCCAGCGGGGACTTCTGCAACAGCTGCAGC-----TGCACAACTGTG 998

DB 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnAsnCysAsnAsnLeu 120

QY 999 CGGCATGAGCTCGAGCGCTTCAAGCCATAAGCGCTGTCTGTATAGAAATCTGAAGCT 1058

DB 121 HisHisAspIleGluArgPheLysAlaIleLysAlaCysLeuGlyArgAsnProGluAla 140

QY 1059 TTCCCAACCAAAATGGGAAGCGCTCTGGGAGCTGCTAACTTCGACACAGCAAGGG 1118

DB 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160

QY 1119 TGCAGCTGAAGCGCTCAGCTGCCTGAAGAACTACTGTGAGTGTATGAGCGCAAAATC 1178

DB 161 CysAsnCysArgArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGlnIle 180

QY 1179 ATGTGTCTTCCATTTGCCAAATTCATTTGCTTGCAGAACTATGCAAAAGTCCAGACGA 1238

DB 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200

QY 1239 AAAATCTGATGAGCACACCCACTACATGAGCGCTGGGACCTTTCAGAGCAGCCATTAT 1298

DB 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220

QY 1299 TTGTCGCCAGCCAAAGTTCTCAGGACCTCCAAACTGAGAAAATAGGAGCGCTTCTCC 1358

DB 221 LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgArgProSerSer 240

QY 1359 TGTATCTCTGGGAAGTAGTGGAGGCCACATGTGCTGCTGCTGCCCGAGGTGAGGAA 1418

DB 241 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 260

QY 1419 GCAGAGCAGGAGCAGCTGTTCCCAAGCTGGCTGAGCAGATGATCTGAGGAGGTTTGGGA 1478

DB 261 AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluPheGly 280

QY 1479 AGGTGCTGTGCGAGATTTCTCCACATCGAGTTCACATGCCAAGGGGCTGAAATTTAG 1535

DB 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 3

ABG74678

ID -ABG74678 standard; Protein: 528 AA.

XX

AC ABG74678;

XX

DT 10-MAY-2003 (first entry)

XX

DE Human CGDD protein 2369279CD1 SEQ ID 4.

Db	248	GlyTyrAlaValLeuProAlaGlnTyrValThrGlnLeuGlnInlSerSerTyrValSer	267	XX	Homo sapiens.
QY	873	GCTGTCAATGGCGCTTCCCTCTGGACCTGCTGCAAGGCCACCCAAAATAACT	932	XX	WO200155317-A2.
Db	268	IleAlaSerAsnSerThrPheThrGlyThrSerGlyIleGlnThrGlnAlaArgLeuPro	287	XX	02-AUG-2001.
QY	933	CTGCTCTGGG-----	941	XX	17-JAN-2001; 2001WO-US01329.
Db	288	PheAsnGlyIleIleProSerGluSerAlaSerArgProArgLysProCysAsnCysThr	307	XX	31-JAN-2000; 2000US-0179065.
QY	942	-----TACTGTGACTCTCTCCAGCGGGGACTTCTGCAACAGC	980	PR	04-FEB-2000; 2000US-0180628.
Db	308	LysSerLeuCysLeuLysLeuTyrCysAspCysPheAlaAsnGlyGluPheCysAsnAsn	327	PR	24-FEB-2000; 2000US-0184664.
QY	981	TCAGCTGC-----AACCACTGCGCCCTAGCTGAGCTTCAAGCCATA	1028	PR	02-MAR-2000; 2000US-0186350.
Db	328	CysAsnCysThrAsnCysTyrAsnAsnLeuGluHisGluAsnGluArgGlnLysAlaIle	347	PR	16-MAR-2000; 2000US-0189874.
QY	1029	AAGCGTGTCTGTATAGAAATCCTGAAGCTTTCCAAACCAAAATGGGAAAGCGCTGTG	1088	PR	17-MAR-2000; 2000US-0190076.
Db	348	LysAlaCysLeuAspArgAsnProGluAlaPheLysProLysIleGlyLysGlyLysGlu	367	PR	18-APR-2000; 2000US-0198123.
QY	1089	GGAGCTGTAACCTCGACAGCAAGAGGTCCAACTGTAAGCGCTCAGGCTGCTGAAG	1148	PR	19-MAY-2000; 2000US-0205515.
Db	368	GlyGluSerAspArgHisSerLysGlyCysAsnCysLysArgSerGlyCysLeuLys	387	PR	07-JUN-2000; 2000US-0209467.
QY	1149	AACTACTGTGAGTGTATGAGCGCCAAATCATGTCTTCCATTTGCAAAATGCATTGCT	1208	PR	30-JUN-2000; 2000US-0214886.
Db	388	AsnTyrCysGluCysTyrGluAlaLysIleMetCysSerSerIleCysLysCysIleGly	407	PR	07-JUL-2000; 2000US-0215135.
QY	1209	TGCAAAATATGAAGAAAGTCCAGAACCAAAATGCTGATGAGCACACCCCACTACATG	1268	PR	07-JUL-2000; 2000US-0216647.
Db	408	CysLysAsnPheGluSerProGluArgLysThrLeuMet-----HisLeuAla	424	PR	11-JUL-2000; 2000US-0217487.
QY	1269	GAGCTGGGGACTTT-----GAG	1286	PR	14-JUL-2000; 2000US-0217496.
Db	425	AspAlaAlaGluValArgValGlnGlnThrAlaAlaLysThrLysLeuSerSerGln	444	PR	14-JUL-2000; 2000US-0218290.
QY	1287	ACGACCAATTATTG-----TCCCGACGCAAGTTCTCAGGACCTCCAAACTGAGA	1337	PR	26-JUL-2000; 2000US-0220963.
Db	445	IleSerAspLeuLeuThrArgProThrProAlaLeuAsnSerGlyGlyLysLeu---	463	PR	14-AUG-2000; 2000US-0220964.
QY	1338	AAAAATAGCAGCGCTTCTCTGTATCTCTGGGAAGTAGTGAGGCGCCACATGTCCTGC	1397	PR	14-AUG-2000; 2000US-0224518.
Db	464	-----ProPheThrPheValThrLysGluValAlaAlaThrCysAsnCys	479	PR	14-AUG-2000; 2000US-0224519.
QY	1398	CTGCTGGCCCGGTGAGGAACAGCAGGAGCACTGTTCCCAAGCTTGCTGCTGAGCAG	1457	PR	14-AUG-2000; 2000US-0225213.
Db	480	LeuLeuAlaGlnAlaGlnAlaAspLysLysGlyLysSerLysAlaAlaAlaGluArg	499	PR	14-AUG-2000; 2000US-0225214.
QY	1458	ATGATCCTGAGGAGTTTGAAGGTGCTGTCGCGAGATTCTCCACATCGAGTCAAGTCC	1517	PR	14-AUG-2000; 2000US-0225267.
Db	500	MetIleLeuGluGluPheGlyArgCysLeuMetSerValIleAsnSerAlaGlyLysAla	519	PR	14-AUG-2000; 2000US-0225268.
QY	1518	AAG 1520		PR	14-AUG-2000; 2000US-0225270.
Db	520	Lys 520		PR	14-AUG-2000; 2000US-0225447.
RESULT 4					14-AUG-2000; 2000US-0225757.
ABB96025					14-AUG-2000; 2000US-0225758.
ID	ABB96025 standard; Protein; 147 AA.				14-AUG-2000; 2000US-0225759.
XX					18-AUG-2000; 2000US-0226279.
AC					22-AUG-2000; 2000US-0226581.
XX					22-AUG-2000; 2000US-0226586.
DT	21-JUN-2002 (first entry)				22-AUG-2000; 2000US-0227182.
XX					23-AUG-2000; 2000US-0227009.
DE	Human testicular antigen SEQ ID NO: 1409.				30-AUG-2000; 2000US-0228924.
KW	Human; testicular antigen; testes; cancer; metastasis; immune disorder;				01-SEP-2000; 2000US-0229287.
KW	reproductive system disorder; urinary system disorder; gene therapy;				01-SEP-2000; 2000US-0229343.
KW	cardiovascular disorder; respiratory disorder; neurological disorder;				01-SEP-2000; 2000US-0229344.
KW	gastrointestinal disease; infection; cytostatic.				01-SEP-2000; 2000US-0229345.

PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 DR N-PSDB; AAL01300.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 XX
 PS Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a protein of the invention.
 XX
 SQ Sequence 147 AA;
 Alignment Scores:
 Pred. No.: 1.31e-38 Length: 147
 Score: 503.00 Matches: 97
 Percent Similarity: 80.15% Conservative: 12
 Best Local Similarity: 71.32% Mismatches: 27
 Query Match: 12.55% Indels: 0
 DB: 22 Gaps: 0
 US-09-743-237-1 (1-2241) x AAB95330 (1-147)
 QY 495 GTTCAATCAAGAGCAGGTGGTAGTGTGCCAGGCGCAGCTTTCAGAGCGAGCTTTC 554
 DB 8 ValGluLeuLysGluAlaGlyGlyThrThrSerAsnAsnProGluGluAlaThrLeu 27
 QY 555 CAGGCCCTCTGGCTCAGGAATCTGTGTGCAAGTTCATCCAGGAGGCGAGAGGAG 614
 DB 28 GlnAsnLeuLeuAlaGlnGluSerCysCysLysPheProSerSerGlnGluLeuAsp 47
 QY 615 GCCTCCAGCTCCCTCGGAAGAAGACTCCAGCCCGATGGTGTTCAGCTGAAAGGA 674
 DB 48 AlaSerCysCysSerLeuLysLysAspSerAsnProMetValIleCysGlnLeuLysGly 67
 QY 675 GCGGCCAGATGCTCTGCATAGACAACCTGTGGCGGAGGAGCTCAAGCGCTCCATCTG 734
 DB 68 GlyThrGlnMetLeuCysIleAspAsnSerArgThrArgGluLeuLysAlaLeuHisLeu 87
 QY 735 CTTCCTCAGTACGATGACGAGCAGGTTCCTCCTCAGTACGAGCTCCCTTAAGCCCAATGACA 794
 DB 88 ValProGlnTyrGlnAspGlnAsnTyrLeuGlnSerAspValProLysProMetThr 107
 QY 795 ACTTAGTGGGAAGACTTCTGCCAGTACCGAGGAAGTTAAATCATCATCACACAGGTGAT 854
 DB 108 AlaLeuValGlyArgPheLeuProAlaSerThrLysLeuAsnLeuIleThrGlnGlnLeu 127
 QY 855 AATGGAGCTCTCCATCAGCTGTCAATGGGCTGCTTCCCTCTGGA 902
 DB 128 GluGlyAlaLeuProSerValValAsnGlySerAlaPheProSergly 143
 RESULT 6
 ABB62035
 XX
 AC ABB62035 standard; Protein; 950 AA.
 XX
 AC ABB62035;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 12897.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO2001/1042-A2.
 XX
 PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL06138.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 12897; 2lpp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 950 AA;
 Alignment Scores:
 Pred. No.: 2.64e-32 Length: 950
 Score: 440.50 Matches: 138
 Percent Similarity: 45.87% Conservative: 62
 Best Local Similarity: 31.65% Mismatches: 141
 Query Match: 10.99% Indels: 95
 DB: 22 Gaps: 14
 US-09-743-237-1 (1-2241) x ABB62035 (1-950)
 QY 257 CCCCTCTGTAAGTAAGGACAGTCTTTAAACCTATCTAGAGTTTTCATGAATTCGTCT 316
 DB 572 ProValGlnLysLeuIleAsnArgThrAlaValGlnArgValValSerSerSerThr 591
 QY 317 TCTCAACCATATTGCTAAGCTATATAGCAATTCCTTGAATTCCTATATACTAGGAGA 376
 DB 592 SerPro-Ser-----SerAsnSerThrLysLysIleTyrAsnTyrValGln 606
 QY 377 ACCTCTGATTCCTGCTCCTCTACATCTCCTGAGTGTAGTGTACAGGGGAAATCATTTTG 436
 DB 606 nPro-----
 QY 437 GTGAGACTCCGATGAACCTACTCCAGGTTCCAGGAGCAGCAAGCAAGCAAGAAAGTGT 496
 DB 608 -----ThrGlySerAsnGlyAlaLysTyrMetIleCysAsnSe 620
 QY 497 TGAATCAAGAAGCAGGTGGTAGTGTGCCA-----GGCGCAGCCCTGGAAGA 544
 DB 620 rGlyValProGlnSerSerThrSerAlaMetArgGlyTyrThrGlyTyrValGluAs 640
 QY 545 CGCAGCTTCCAGCCCTCTGGCTCAGGAATCTGTTGCAAGTTCCTCATCATCCAGGAGA 604
 DB 640 n-LysThrArgArgPro-----ProIleSerProGlnGlnHisArgPheL 656
 QY 605 GGCAGAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCTGATGTTGTCA 664
 DB 656 ysGlnMetGlyPro-GlnGlnGlnSerLys-----HisGlnGlnLeuGln 670
 QY 665 GCTGAAAGGAGGCGCCAGATGCTCTGCTATAGACAACCTGTGGCGCAGGAGCTCAAGC 724
 DB 665 GCTGAAAGGAGGCGCCAGATGCTCTGCTATAGACAACCTGTGGCGCAGGAGCTCAAGC 724

Db 671 AlaGlnAlaLysGln-----ArgIleArgGlnGlnLeuProThrGluGlnSer 687
 QY 725 GCTCCATCTGCTTCCCTCAGTACGACGAGCAGTTCCTCCAGTCAGAGCTCCCTAA 784
 Db 688 ThrProIleLysValGluProLysLeuProThrLeuProGlyValLysAla-AsnVa 707
 QY 785 GCAATGACAACTTTAGTGGGAAGACTTCTGCGCAGTACGAGCAAGTTAAATCTCATC 844
 Db 707 lProAlaLysProLeuPheGluValLeuLysPro--ProAla----- 720
 QY 845 ACAGGTGATAGTGGAGCTCCCATCAGCTGCAATGGGCGCTTTCCTCTCTGACC 904
 Db 721. -----ThrAlaAlaAlaGlyAlaValAspProLeuGlyG 733
 QY 905 TGCTCTGCAAGGCCACCAATAAATCTGTCTGGG-----TACTG 946
 Db 733 yMetThrSerArgArgLysHisCysAsnSerLysSerGlnCysLeuLysLeuTyrCy 753
 QY 947 TGACTGCTTCTCAGCGGGGACTTCTGCAACAGCTCGACTGC-----AACAA 994
 Db 753 sAspCysPheAlaAsnGlyGluPheCysGlnAspCysThrCysLysAspCysPheAsnAs 773
 QY 995 CCTGCGCCATGACTCGAGCGCTTCAAGCCATAAAGCGGTGCTTGATAGAATCCTGA 1054
 Db 773 nLeuAspTyrGluValGluArgAlaIleArgSerCysLeuAspArgAsnProSe 793
 QY 1055 AGCTTTCCAAACAAAATGGGAAAGCGCTCTGGGAGCTGCTTAAACTTCGACACAGCAA 1114
 Db 793 rAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetArgLeu---HisAsnLy 812
 QY 1115 AGGTCGCACTGAAGCGCTCAGCGCTGCTGGAAGACTTCTGAGTGTGCTATGAGCCAA 1174
 Db 812 sGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLy 832
 QY 1175 AATCATGTGTCTTCCATTGCAATGCAATGCTTGCATAAACTATGAAGAAAGTCCAGA 1234
 Db 832 sIleProCysSerSerIleCysLysCysValGlyCysArgAsnMetGluAspArgProAs 852
 QY 1235 ACGAAATAATGCTGATGAGCACACCCCACTACATGAGGAGCTGGGAGCTTTGAGAGAGCCA 1294
 Db 852 p-----ValAspMetAspSerLeuAs 859
 QY 1295 TTATTGTGCCCCAGCAAGTCTCAGGACCTCCAAACTCAGAAATAATAGGCAG----- 1349
 Db 859 pGlyLeuMetGlyValGluGlyGlnLysLysAspLys---AlaLysAsnLysGlnLeuAs 878
 QY 1350 -----GCCTTCTCCTGTATCTCCTGGGAAGTAGTGGAGGCCACATGTGCTGCTC 1399
 Db 878 nGluAsnArgAlaAsnIleTyrPheThrAspValIleGluAlaThrIleMetCysMe 898
 QY 1400 GCTGGCCCGAGGTGAGAGCAGCAGCAGGAGCAGCTGCTCCCAAGCTTGGCTGAGCAGAT 1459
 Db 898 tIleSerArgIleValMetHisGlyLysGlnAsnValAlaValGluAspMetGluArgG 918
 QY 1460 GATCCTGGAGGAGCTTTGGAGGTGCTGTCGCGAGATTCTC 1499
 Db 918 uValMetGluGluMetGlyLysSerLeuThrGlnIleIle 931

RESULT 7

AAW83392

ID AAW83392 standard; Protein; 438 AA.

AC AAW83392;

XX 29-MAR-1999 (first entry)

DT Caenorhabditis elegans synMuv protein LIN-54.

DE LIN-54; synthetic multivulvar; synMuv; signal transduction;

XX animal model; tumour suppressor; retinoblastoma; cancer; cancer;

KW cell proliferation; gene therapy.

XX Caenorhabditis elegans.

OS

XX WO9854299-A1.
 PN 03-DEC-1998.
 PD 28-MAY-1998; 98WO-US11043.
 PF 28-MAY-1997; 97US-0047996.
 PR (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PA Ceol C, Horvitz HR, Lu X;
 PI WPI; 1999-045362/04.
 DR N-PSDB; AAV72865.
 DR Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
 PT useful for treating diseases associated with altered levels of cell
 PT proliferation, e.g. carcinomas
 XX Claim 7; Fig 13; 70pp; English.
 PS This is the amino acid sequence of LIN-54, a novel protein of
 CC Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
 CC synthetic multivulvar (synMuv) gene involved in cell fate and cell
 CC proliferation, and is part of a pathway that may be used as a
 CC genetic and biochemical model system for tumour suppression and
 CC cancer in mammals. synMuv pathway genes and proteins may be used
 CC to identify genes which are part of the mammalian pathway and
 CC to identify genes, proteins and therapeutic compounds which
 CC modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
 CC C. elegans synMuv polypeptides selected from LIN-37, -35, -55, -52,
 CC -53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
 CC (1) nucleic acids (see AAV72866-67) encoding the mouse and human
 CC homologues of C. elegans LIN-54; (2) vectors containing the nucleic
 CC acids; (3) transgenic cells; (4) a pure mammalian synMuv family
 CC polypeptide, and (5) an antibody which binds to a synMuv family
 CC protein. The synMuv nucleic acids and polypeptides can be used to
 CC diagnose and treat, especially by gene therapy, conditions
 CC involving altered levels of cell proliferation, e.g. synMuv-
 CC associated carcinomas.
 XX SQ Sequence 438 AA;

Alignment Scores:

Pred. No.: 3,77e-29 Length: 438
 Score: 405.50 Matches: 132
 Percent Similarity: 44.37% Conservative: 57
 Best Local Similarity: 30.99% Mismatches: 158
 Query Match: 10.12% Indels: 79
 DB: 20 Gaps: 18

US-09-743-237-1 (1-2241) x AAW83392 (1-438)

QY 419 CAGGGGGAATCATTTTGTGTGAGACTCCGATGAACCTGCCAGGTGCCAAGCAGCAA 478
 Db 3 Gnglygluileval--TyrGlnAspAspAspTyrTyrAspGluSerGluIleTyrAs 22
 QY 479 GCAAGCAAGAAAAAGTGTGTAATCAAGAACAGGTGTGTAGTGCCAGCGCGCAGCCC 538
 Db 22 pAsnTyrGluGluGlyAlaGluPheIleGluValAsnGlyGlnLeuValProHisAsnPr 42
 QY 539 TGAAGACGCGAGCTTTCAGGCGCCCTCTGGCTCAGGAATCCTGTGTGAAGTTCCTCATC 598
 Db 42 o-AsnLeuGlnAlaGlnGln-----AsnArgProGlyThrSerMetIleG 58
 QY 599 CCAGGAGCGCAGAGGAGCGCTCCAGCTGCCCTCGAAGAAAGACTCCAGCGCCCATGCTGAT 658
 Db 58 lncIlnHisAsnArgSerMetGluValAsnGlnGlyLeuValLysAspGluProIle-Asp 77
 QY 659 TTGTGAGCTGAAAGG-----AGGCGCCCGCAGATGCTCTGCATAGACAACCTGTGCGCG 709
 Db 78 ThrSerSerHisArgValTyrValProProArgProValGlnArgLysLeuTyrLys 97

134 AsnAsnLeuGluHisGluAsnGluArgGlnLysAlaIleLysAlaCysLeuAspArgAsn 153
 1050 CCTGAAGCTTCAACCAAAATGGGAAAGCGCTCTGGAGCTGCTAAACTTCGACAC 1109
 154 ProGluAlaPheLysProLysIleGlyLysGlyLysGluGlyLysSerAspArgArgHis 173
 1110 AGCAAGGTGCAACTGTAAAGCGCTCAGGCTGCCTGAAGAATCTAGTGTCTATGAG 1169
 174 SerLysGlyCysAsnLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGlu 193
 1170 GCCAAATCATGTCTTCATTCCTCAATGCAATGCTGCTGCAAAACTATGAAGAACT 1229
 194 AsnIleIlePheIleTrpTyrCysCysArg-----TyrArgGlnIle 207
 1230 CCAGAACGAAAAATGCTGATGAGCACACCCCACTACATGAG 1271
 208 LeuGluLysAlaIleGlnLeuSerGlyAlaGluGlnLeuGlu 221

RESULT 9
 AAM93348
 ID AAM93348 standard; Protein; 251 AA.

AC AAM93348;
 XX
 XX 06-NOV-2001 (first entry)
 XX Human polypeptide, SEQ ID NO: 2895.

DE Human; full length cDNA; cDNA synthesis; oligo-capping.

KW Homo sapiens.

OS EP1130094-A2.

PN 05-SEP-2001.

PD 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

PA (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

DR N-PSDB; AAK94268.

XX 830 Primers useful for synthesizing full length cDNA clones and their

PS use in genetic manipulation -

XX Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 251 AA;

Alignment Scores:

Pred. No.: 1.25e-23 Length: 251

Score: 346.00 Matches: 66
 Percent Similarity: 80.65% Conservative: 9
 Best Local Similarity: 70.97% Mismatches: 18
 Query Match: 8.63% Indels: 0
 DB: 22 Gaps: 0
 US-09-743-237-1 (1-2241) x AAM93348 (1-251)
 QY 495 GTTGAATCAAAAGACAGGTGTGTGTCAGGCGCGCAGCCTGAAGACGAGCTTTC 554
 158 ValGluIleLysGluAlaGlyGlyThrThrThrSerAsnAsnProGluGluAlaThrLeu 177
 QY 555 CAGGCCCTCTGGCTGAGAAATCCCTGTCAGATCCCATCTCCAGGAGGAGAGAG 614
 178 GlnAsnLeuAlaGlnGluSerCysLysPheProSerSerGlnGluLeuGluAsp 197
 QY 615 GCCTCCAGCTGCCTCGGGAAGAAAGACTCCAGCCCATGTTGTCAGCTGAAGGA 674
 198 AlaSerCysSerLeuLysAspSerAsnProMetValIleCysGlnLeuLysGly 217
 QY 675 GCGGCCAGATGCTGTGCATAGACAACTGTGGCGCGAGGAGCTCAAAGCGTCCATCTG 734
 218 GlyThrGlnMetLeuCysIleAspAsnSerArgThrArgGluLeuLysAlaLeuHisLeu 237
 QY 735 CTTCTCAGTACGATGACAGCAGCAGCTTTCCTCAGTCA 773
 238 ValProGlnTyrGlnAspGlnAsnTyrLeuGlnSer 250

RESULT 10

AAG76158

ID AAG76158 standard; Protein; 53 AA.

AC AAG76158;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:6922.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 11.

OS Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

DR N-PSDB; AAK35563.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 8365; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing

QY 1427 GGAGCACTGTTCCC 1440
DB 498 euAlaGlnLeupro 502
RESULT 12
ABB68888
ID ABB68888 standard; Protein; 243 AA.
XX ABB68888;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33456.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12991.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure: SEQ ID NO 33456; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 243 AA;
Alignment Scores:
Pred. No.: 1.74e-08 Length: 243
Score: 186.00 Matches: 59
Percent Similarity: 36.64% Conservative: 26
Best Local Similarity: 25.43% Mismatches: 61
Query Match: 4.64% Indels: 86
DB: 22 Gaps: 8
US-09-743-237-1 (1-2241) x ABB68888 (1-243)
QY 1044 AGAATCTCGAGCTTTCACCAAAATGGGAAGGCGCTCTGGAGCTGCTAACTT 1103
DB 7 ArgSerValAspLysAlaAspGlyLysLysGlyGln---GlyAlaGlyGlyVal 25
QY 1104 CGACACAGCAAGGGTGAACCTGTAAAGCGCTCAGCGCTCGAAGAACTACTGTGAGTGC 1163
DB 26 -----LysGlyCysCysLysArgSerGlnCnSileIeyAsnTyrCysAspCys 42

QY 1164 TATGAGGCCAAATCATGTGTTCTTCCATTTCCTCAATGCTTGTGCAAAACATGAA 1223
DB 43 TyrGlnSerMetAlaIleCysThrLysPheCysArgCysValGlyCysArgAsnThrGlu 62
QY 1224 -----GAAAGTCCAGACGA--- 1238
DB 63 ValArgGluLeuValAspProAsnSerValAlaLysAsnSerSerAlaValLysArgGln 82
QY 1238 ----- 1238
DB 83 LysAlaAlaAlaMetSerAlaLysAlaAlaAlaAlaAlaLysAlaGlyIleAspVal 102
QY 1239 -----AAAATGCTGATG 1250
DB 103 GlnGlyLysAlaLeuGlnValAlaAlaSerThrLeuAlaLeuProGlyLysAlaLeuMet 122
QY 1251 AGCACACCCACTACATG-----GAGCCTGGGGACTTTGAGAGCAGCCATTTAT 1298
DB 123 ThrProLysTyrThrLeuValAlaGlyLysPro---ProMetAlaSerSerHis--- 140
QY 1299 TTGTCCCCAGCCCAAGTTCTCAGGACCT----- 1325
DB 141 IleAsnProIleProIleSerArgProIleAlaThrAlaAlaThrProAlaArgAlaVal 160
QY 1326 -----CCAAACTGAGAAAAAT 1343
DB 161 LysGlnProAlaGluProProMetProValAsnLeuIleIleProValArgHisAspAsp 180
QY 1344 AGGAGGCGCTTCTCTGTATCTCTGGGAAGTAGTGGAGGCCACATGTCCTGCCTGCTG 1403
DB 181 ArgArgAspArgAsnLeuPheValGlnProValAsnAlaAlaLeuLeuGluCysMetLeu 200
QY 1404 GCCCAGGGTGAAGAACGAGCAGCAGCAGCTGTCCTCCCAAGCTTGCTGAGCAGATGATC 1463
DB 201 IleGlnAlaThrGluAlaGluGlnLeuGlyLeuAsnGluLeuGlnValCysGlnLeuVal 220
QY 1464 CTGAGGAGTTTGAAGGCTGCTGTCGCGAGATTCTC 1499
DB 221 LeuGluGluPheMetArgGlyTyrLysAsnIleLeu 232
RESULT 13
ABG30203
ID ABG30203 standard; Protein; 4561 AA.
XX
AC ABG30203;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #30194.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS94390.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

516	Qy	GTGAGTGTGCCAGGGCGAGCCCTGAAGACGCAGCTTTCAGGCGCCCTCTGGCTCAGGAA	575
1772	Db	GLYThAlaProProAlaAlaPheGlyGlyAlaGluCysGlnProThrMetGluAla	1791
576	Qy	TCCTGTTGCAAGTTC-----CATCATCCAGGAGGCAGAGGAGGCGCTCCAGCTGC	626
1792	Db	GLuPheCysSerLeuArgProCysProAlaSerTrpGlyCysCysHisArgValProCys	1811
627	Qy	CCTCGGAAGNAGACTCCAGCCCGCATGGTGATTTGTCAGCTGAAGAGGCGGCCAGATG	686
1812	Db	ThrGlyGluLeuGluGlnArgProLeuMetValSerGlnIle-----	1825
687	Qy	CTCTGCATACACAACGTGTGGCGGAGGAGCTCAAAGCGCTCCATCTGCTTCTCAGTAC	746
1826	Db	-----LeuGluAla-----	1828
747	Qy	GATGACCAGACAGTATTTCCCTCAGTCAGAGCTCCCTAAGCCATGACAAC	800
1829	Db	GlnAspGlnGlyValAlaProValSerProGlySerSerProThrAlaAlaProGlyLeu	1848
801	Qy	GTGGGAGACTTCTGCACA-----GTACCAGCGAAGTTAAATCTCATCACACAG-----	848
1849	Db	ValArgArgValSerProArgGlyHisProGlyProLeuGlyAlaSerValGlnProPro	1868
849	Qy	-----GTTGAATATGGAGCTCTCCCATAGCTGTCAATGGGGTGC	890
1869	Db	ValAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAla-----GlyGlyTrp	1886
891	Qy	TTTCCCTCTCGACCT-----	905
1887	Db	GlyProTrpGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArg	1906
906	Qy	-----GCTCTGCAAGGCGCCACCAAAATAACTCTGTCTGGGTACTGTGAC--TGCTTC	956


```

Qy 1062 CAACAAAA-----ATGGGAAAGCCGCTG 1088
Db      |||
2701 CysProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnProGlyHisCys 2720
Qy 1089 GGAGCTCCTAACTT-----CGACACAGCAAGGGTCAACTGTAAAGCGC--- 1133
Db      |||
2721 SerCysLeuAspLeuLeuThrGlyGlnArgHisProGlyAlaArgLeuAlaArgPro 2740
Qy 1134 TCAGGCTGCCTGAAGAACTACTGTGAGTGTATGAGGCCAAATCATGTCTTCTCCATT 1193
Db      |||
2741 AspGlyCys-----AsnHisCysThrCysLeuGluGlyArgLeuAsnCysThrAspLeu 2758
Qy 1194 TGCAAAATGCATTCCTCCAAAACATATGAAGAAAGTCCAGAACGAAAAATGCTGATGAGC 1253
Db      |||
2759 ---ProCysProAspCys----- 2763
Qy 1254 ACACCCCACTACATGGAGCCTGGGACTTTGAGAGCAGCCATATTTGTCCCCAGCCAAAG 1313
Db      |||
2764 -----GlyGlyGlnSerLeuHis-ProCysGlyGlnProCys 2776
Qy 1314 TTCTCAGGACCTCCAAAACGTAGAAAAAATAGCGAGCCTCTCTCTGTATCTCTCTGGAA 1373
Db      |||
2776 sProArgSerCysGlnasp-----LeuSerProGlySe 2787
Qy 1374 GTAGTGGAGGCCACATGTGCTGCTGCTGCTGCCAGGCTGAGGAAGCAGAGAGCAC 1433
Db      |||
2787 rValCysGlnProGlySerValGlyCysGlnPro-----ThrCysG1 2801
Qy 1434 TGTTCCTCCAGCTGGCTGAGCAGATCATCTGAGGAGTTTGGAAAGTGCCTGTCTCCAG 1493
Db      |||
2801 yCysProLeuGlyGlnLeuSerGln-----AspGlyLeuCys----- 2813
Qy 1494 ATTCTCCACATCGAGTTCAAGTCCAGGGGCTGAAATTTGAGTAGCTGCAAGCTGTAA 1553
Db      |||
2814 -----ValProProAlaHi 2818
Qy 1554 AGGGAAATGCTGTGGCAAGCCTCAGCCCTGGGAATCTGCACGAGGAAGCTGTGCCCA 1613
Db      |||
2818 sCysArgCysGlnTyrGlnProGlyAlaMetGlyIleProGluAsnGlnSerArgSerAl 2838
Qy 1614 GGGAGGAGCAGAGCCCGCATCATG-----CCAGTCACTGCTGAGTCTG 1661
Db      |||
2838 aGly--SerArg-----PheSerSerTrpGluSerLeuGluProGlyGluVal----- 2853
Qy 1662 AGTCATCTGCTGCTACTGG-----CAGCCTACTCAAGGTATCTCTAAAG 1706
Db      |||
2854 -----ValThrGlyProCysAspAsnCysThrCysValAlaGlyIleLeuGln 2869
Qy 1707 TGCAAGCAGGCA-----GAGCCA-----CCTGGGGA---TGGACA 1739
Db      |||
2870 CysGlnGluValProAspCysProAspProGlyValTrpSerSerTrpGlyProTrpGlu 2889
Qy 1740 CTGGCCCTCTCTCCCTGGGAGCCCTCTGGGACNCCCTG----- 1781
Db      |||
2890 AspCysSerValSerCysGly-----GlyGlyGluGlnLeuArgSerArgCysAla 2907
Qy 1782 -----CCTGCATAAAGAGGGTGATTTTCTACTTGTGTATGTGTTTCTTCAAA 1835
Db      |||
2908 ArgProProCysProGlyProAlaArgGlnSerArgThrCys----- 2921
Qy 1836 TTGCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCAGCCTCAAGTTAGAGACTAGGC 1895
Db      |||
2922 -----SerThrGlnVal-----CysArgGluAlaGly 2930
Qy 1896 TCTTCTTCAGTGGACTC-----TGCCCAATACATACAGTCAGTGGCCATCA 1946
Db      |||
2931 CysProAlaGlyArgLeuTyrArgGluCysGln----- 2941
Qy 1947 GGGGTTTTCCAGCCAGGCTGTGCACAGGAGATATGGAGGGGGTGGTTAGAGCTG 2006
Db      |||
2942 -----ProGlyGluGlyCysProPheSerCysAlaHis-----ValThrGlnGlnVal 2957
Qy 2007 GGTTTGTGTTGATTTTGGGTTTTTTTCTCTCTGTTATTTCTGCTTGAAGTGAAGAACT 2066

```

```

Db 2958 Gly-----CysPheSerGluGlyCysGluGluGly 2967
Qy 2067 TGCTCTCTGTCCA-----ACCTTTTCTCCATATTAATTACTGCTGCACGGTCCCTGCTGAC 2120
Db 2968 Cys-HisCysProGluGlyThrPheGln-HisArgLeuAlaCysValGlnGluCys---P 2986
Qy 2121 CAGTCACAGTGACCTCA 2137
Db 2986 rOCysValLeuThrAla 2991

RESULT 15
AAU80189
ID AAU80189 standard; Protein; 966 AA.
XX AC AAU80189;
XX DT 15-JUL-2002 (first entry)
XX DE Human TSPI domain containing protein encoded by cDNA FG01869.
XX KW TSPI; thrombospondin domain; FG01869; angiogenesis; vasculogenesis; human.
XX OS Homo sapiens.
XX PN JP2002085059-A.
XX PD 26-MAR-2002.
XX PF 08-SEP-2000; 2000JP-0273778.
XX PR 08-SEP-2000; 2000JP-0273778.
XX PA (KAZU-) 2H KAZUSA DNA KENKYUSHO.
XX PA (YOSH ) YOSHITOMI PHARM IND KK.
XX DR WPI; 2002-378268/41.
XX PT N-PSDB; ABK50390.
XX PS TSPI domain-containing polypeptide useful for drug compositions -
XX Claim 1; Page 30-33; 51pp; Japanese.

The invention relates to a TSPI (thrombospondin 1) domain-containing polypeptide comprising the proteins appearing as AAU80189 and AAU80189, encoded by cDNAs designated FGO6969 and FG01896. Also included are proteins that are 50% homologous to the proteins and a polypeptide having at least one deletion, replacement, addition or insertion of amino acid in the proteins and having at least 8 repetitions of the TSPI domain. The polypeptide can be used in drug compositions particularly for disorders associated with angiogenesis and vasculogenesis. The present sequence is the TSPI domain containing protein encoded by cDNA FG01869.

SQ Sequence 966 AA;

Alignment Scores:
Pred. NO.: 0.00164 Length: 966
Score: 136.50 Matches: 120
Percent Similarity: 29.83% Conservative: 53
Best Local Similarity: 20.69% Mismatches: 204
Query Match: 3.41% Indels: 203
DB: Gaps: 29

US-09-743-237-1 (1-2241) x AAU80189 (1-966)
Qy 435 TGGTGAGACTCCGATGAACACTACTGCCAGGTTCCTCAAGCAGCAGCAAGCAAGAAAGT 494
Db 237 TrpSerTrpCysAspArgSerCysGlyGlyGlnSerLeuArgSerArgSerCysSer 256
Qy 495 GTTGAATCAAGAACGAGGTGGTAGTGTGCCGCGCGCCCTGGAAGACGCGCTTC 554

```

Db 257 SerProProSerLysAsnGlyGlyAlaProCysAlaGlyGlyArgHisGlnAlaArgLeu 276
 QY 555 CAGGCCCTCTGCTCAGGAATCCTGTGCAAGTTCCTCATCCAGGAGGAGGAG 614
 Db 277 CysAsnProMetProCysGluAlaGlyCys-----ProAlaGlyMetGluValValThr 294
 QY 615 ---GCCTCAGCTGCTCGGAAG---AAAGACTCCAGCCCATGTGATTTGTACGCTG 568
 Db 295 CysAlaAsnArgCysProArgArgCysSerAspLeuGlnGlyIleValCysGln--- 313
 QY 669 AAAGGAGGCCCCAGATGCTCTGCATAGACAACACTGTGGCGCGAGGAGCTCAAAGGCTC 728
 Db 313 ----- 313
 QY 729 CATCTGCTTCTCAGTACGATGACAGCAGAGTTCCTCCCTCAGTACAGAGTCCCTAAGCCA 788
 Db 314 -----AspAspGlnValCysGlnLysGlyCysArgCysProLys--- 326
 QY 789 ATGACAACCTTTAGTGGAGACTTTCGCCAGTACCAGCAAGTTAAATCTCATCACACAG 848
 Db 327 -----GlySerLeu----- 330
 QY 849 GTTGATAATGGAGCTCTCCCATCAGCTGTCAATGGGGTGCCTTTCCCTCTGGAGCTGCT 908
 Db 331 GlnAspGlyCysValPro----- 337
 QY 909 CTGCAAGGGCCACCCAAAATAACTCTGTGGTACTGTGACTGCTTCTCCAGC----- 962
 Db 338 -----IleGlyHisCysAspCysThrAspAlaGlnGly 348
 QY 963 -----GGGACTTCTGCAACAGCTGCAGCTGC----- 989
 Db 349 HisSerTrpAlaProGlySerGlnHisGlnAspAlaCysAsnAsnCysSerCysGlnAla 368
 QY 990 AACAACTCGGCCATCAGCTCAGAGCTTCAAGCCATAAAGCGCTCTTGATAGAAAT 1049
 Db 369 GlyGlnLeuSerCysThrAlaGlnProCysProProProThrHisCysAlaTrpSerHis 388
 QY 1050 CCTGAAGCTTCCAAACCA---AAAATGGGAAAGCGCTGTGGAGCTGCTAAACATCGA 1106
 Db 389 TrpSerAlaTrpSerProCysSerHisSerCysGlyProArgGly-----Gln 404
 QY 1107 CACAGCAAGGGTGCACACTGTAAAGCGCTCAGC-----TGCTGAGAACTACTGTGAG 1160
 Db 405 GlnSerArgPheArgSerCysThrProGluGlyValIleCysGluAspThrGluCysAla 424
 QY 1161 TGCTATGAGGCCAAAATCATGTCTCTCCATTTGCAAAATGC---ATTGCTTGCAAAAAC 1217
 Db 425 ValProGluAlaTrpThrLeuTrpSerSerTrpSerAspCysProValSerCysGlyGly 444
 QY 1218 TATGAAGAAGTCCAGCAAGAAAATGCTGTAGAGCACACCCACTACATGGAGCCTGGG 1277
 Db 445 GlyAsnGlnValArgThrArgAlaCysArgAlaAlaProHisHisArgSerPro--- 463
 QY 1278 GACTTTGAGAGCAGCCATTATTGTCCCGCAGCAAGTTCTCAGGACTCCAAAACGTGAGA 1337
 Db 464 -----ProCysLeuGlyProAspThrGlnThr 472
 QY 1338 AAAAATAGGAGGCC-----TTCTCTGTATCTCTGGGAAGTAGTGGAGGCCACA 1388
 Db 473 ArgGlnGlnProCysProGlyLeuLeuGluAlaCysSerTrpGlyPro-TrpGlyProCys 492
 QY 1389 TGTGCTGCTGCTGCTGCGCCAGGTGAGGAGCAGACAGCAGGACTGTTCCTCCCAAGCTTG 1448
 Db 492 sSerArgSerCys-GlyProGlyLeuAlaSerArgSerGlySer-Cys---ProCysLeu 510
 QY 1449 GCTGACCATGATCTGAGGAGTTTGGAGGTGCTCTCCAGATTTCTCCACATCGAG 1508
 Db 511 -----MetalAlaAlaAspProThrCysAsnSerThrPheLeuHisLeuAsp 526
 QY 1509 TTCAAGTCCAAAGGGGCTGAAATTTAGTAGCTGCAAGCTGGTAAAGGG----- 1557
 Db 527 Thr-----GlnGlyCysTyr-SerGlyProCysProGly 537

QY 1558 -GAATGCTGTGTGCAAGCC---TCAGCCCTGGGAATCTGCACCGAGGAAGCTGGTCCCA 1613
 Db 537 uGluCysValTrpSerSerTrpSerSerTrpThrArgCysSer----- 551
 QY 1614 GSGAGGAGCAGAGCGCGCCATCATGCGCAGGTCACTGTGAGGTCTGAGTATGATTCGAT 1673
 Db 552 ----- 553
 QY 1674 GGTACTGGCCAGCCTACTCAAGGTATCTTAAAGTGAAGCAGCAGCAGCAGCAGCCTGGGA 1733
 Db 553 gValLeuValGln-----GlnArgTyrArgHisGlnGly-- 564
 QY 1734 TGGACACTGGCCCTCCTGCTCCCTGGGAGGCCCTCTGGGAGCTCCCTGCCCTGCATPAAAA 1793
 Db 565 -----ProAlaSerArgGlyAlaArgAlaGlyAlaProCysThrArg---Le 579
 QY 1794 AGAGGGTGATTTCTACTTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1844
 Db 579 uAspGlyHisPheArg-----ProCysLeuIleSerAsnCysSerGluAspSerCys 596
 QY 1845 AGTACCTCCATTCAGTATTATAGCCAGCCTCAAGTTAGAGAGCTAGGCTCTTCTTCA 1904
 Db 596 sthrProProPheGlu-PheHis-Ala-----CysGlySerProCys 609
 QY 1905 GGTGACTCTGCCCAATCACATACAA----- 1931
 Db 610 AlaGlyLeuCysAlaThrHisLeuSerHisGlnLeuCysGlnAspLeuProProCysGln 629
 QY 1932 -----GTCAGGTGCCCATCAGGGGTTTTCAGGCCAG 1964
 Db 630 ProGlyCysTyrCysProLysGlyLeuLeuGlnAlaGlyCysIleProProGlu 649
 QY 1965 GCCTGTGAC-----AGGAGATATGGAGGGGCTCGGCTTAGAGCTG 2006
 Db 650 GluCysAsnCysTrpHisThrSerAlaAlaGlyAlaGlyMetThrLeu 665

Search completed: July 24, 2003, 13:58:18
 Job time : 151.946 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:48:49 ; Search time 182.427 seconds

(without alignments)
6037.327 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtcaagcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p_model -DEV=xlp
-O=/cpn2.1/USPTO_spool/US09743237/runat_24072003_113104_17672/app_query.fasta_1.4750
-DB=SPTRMBL_23 -QFMT=fastcan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09743237@cgn.1.1.612.@runat_24072003_113104_17672 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%				
No.	Score	Match Length DB ID	Description			
1	1515.5	39.6	475	11	Q8VIE0	Q8VIE0 mus musculus

2	1508.5	39.4	475	11	Q8VIE1	Q8VIE1 mus musculus
3	836.5	21.9	251	4	Q8NCC8	Q8NCC8 homo sapien
4	669.5	17.5	419	11	Q8C107	Q8C107 mus musculus
5	633.5	16.6	277	11	Q9D571	Q9D571 mus musculus
6	483.5	12.6	950	5	Q9V6Q8	Q9V6Q8 drosophila
7	444.5	11.6	435	5	Q95QD7	Q95QD7 caenorhabdi
8	437	11.4	429	5	O62295	O62295 caenorhabdi
9	376	9.8	571	10	Q9SL70	Q9SL70 arabidopsis
10	372	9.7	532	10	Q8GUZ0	Q8GUZ0 arabidopsis
11	370	9.7	536	10	Q8GUZ2	Q8GUZ2 arabidopsis
12	370	9.7	543	10	Q8GUZ1	Q8GUZ1 arabidopsis
13	358	9.4	603	10	Q9SZD1	Q9SZD1 arabidopsis
14	264.5	6.9	601	10	Q9CAV1	Q9CAV1 arabidopsis
15	262.5	6.9	356	10	Q9LW71	Q9LW71 arabidopsis
16	248.5	6.5	896	10	O9ZS22	O9ZS22 glycine max
17	248	6.5	658	10	O23333	O23333 arabidopsis
18	247.5	6.5	609	10	Q9M679	Q9M679 arabidopsis
19	247.5	6.5	609	10	Q8LS48	Q8LS48 arabidopsis
20	243.5	6.4	526	10	Q8H1Q0	Q8H1Q0 arabidopsis
21	243.5	6.4	695	10	Q9LUI3	Q9LUI3 arabidopsis
22	243.5	6.4	695	10	O9LE32	O9LE32 arabidopsis
23	242.5	6.3	526	10	Q94A12	Q94A12 arabidopsis
24	202.5	5.3	593	10	Q9LUI5	Q9LUI5 arabidopsis
25	200	5.2	243	5	Q9VMQ3	Q9VMQ3 drosophila
26	192	5.0	553	10	Q94DS2	Q94DS2 oryza sativ
27	147	3.8	1704	5	Q94446	Q94446 chironomus
28	137	3.6	4123	4	O75851	O75851 homo sapien
29	136.5	3.6	1637	6	Q9XSV8	Q9XSV8 bos taurus
30	136.5	3.6	5146	6	Q8SPM4	Q8SPM4 bos taurus
31	132.5	3.5	1237	11	Q8BKK7	Q8BKK7 mus musculus
32	131.5	3.4	947	11	Q8BKK7	Q8BKK7 mus musculus
33	130.5	3.4	3695	4	Q8TDF8	Q8TDF8 homo sapien
34	130	3.4	2327	13	Q91BG7	Q91BG7 xenopus lae
35	130	3.4	4782	11	Q8K1G6	Q8K1G6 mus musculus
36	128.5	3.4	1145	11	Q8BMN9	Q8BMN9 mus musculus
37	128.5	3.4	1450	11	Q8JZW7	Q8JZW7 mus musculus
38	127.5	3.3	3010	12	Q9DTE1	Q9DTE1 hepatitis c
39	127	3.3	3010	12	Q9DTE7	Q9DTE7 hepatitis c
40	125.5	3.3	5374	10	O99ND0	O99ND0 mus musculus
41	125	3.3	1028	11	O9TLL0	O9TLL0 mus musculus
42	124.5	3.3	1172	10	Q8RVG7	Q8RVG7 arabidopsis
43	124	3.2	969	4	Q96KG6	Q96KG6 homo sapien
44	122.5	3.2	3198	5	Q9U8G8	Q9U8G8 manduca sex
45	122	3.2	3010	12	Q9Q1Y8	Q9Q1Y8 hepatitis c

ALIGNMENTS

RESULT 1
Q8VIE0
ID Q8VIE0 PRELIMINARY; PRT; 475 AA.
AC Q8VIE0:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Tesmin (Hypothetical 50.6 kDa protein).
GN MTL5 OR TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

EMBL: AB057423; BAB64935.1; -;
 DR EMBL: BC024377; AAB24377.1; -;
 DR MGI: 1340029; Mtl5.
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 1.
 KW Hypothetical protein
 SQ SEQUENCE 475 AA; 50615 MW; 4373ED1188DCE79F CRC64;

Alignment Scores:
 Pred. No.: 4,32e-139 Length: 475
 Score: 1515.50 Matches: 290
 Percent Similarity: 78.52% Conservatives: 39
 Best Local Similarity: 69.21% Mismatches: 69
 Query Match: 39.63% Indels: 21
 DB: 11 Gaps: 3

US-09-743-237-3 (1-2134) x Q8VIE0 (1-475)

QY 51 CTCCTCGGGAGTACCCCGGATCCAGAGCTACGCGCTGGAGAGCTGCGCTCTCTG 110
 DB 78 LeuLeuGlyAspPheProGlyLeuProGlyLeuArgSerProAspAlaAala----- 95

QY 111 CAGGCCCGCAGCCCGCCCTGCAACGTGCACCTCTCTCTCTCTCTCTCTCTCTCT 170
 DB 96 -----ProProAlaTyrSerValHisValLeuSerSerLeuLeuProGlyAla 112

QY 171 CGCAGCCCGCGGTGTTTGGCCCTGGGCGCTGGGCTCTGGGAGAGCTCCACCC 230
 DB 113 ArgGlyProAlaLeuLeu-----ProLeu-SerAl 122

QY 231 GGGCGTCCGATGATCCCGAGTTGAATCAAGTAAGCTAGAGTGGTACTACTACAAGTAATA 290
 DB 122 aGlyValArgValIleProValGluIleLysGlu-AlaGlyGlySerValProGlyGlys 142

QY 291 ATCCGGAAGAACCACTTTCGAGATCTTCTGCTCAGCAATCTCTGTCAGTCCCTCCAT 350
 DB 142 erProGlyAspAlaAlaPheGlnAlaProLeuAlaGlnLysSerCysLysPheProS 162

QY 351 GGTCCAGCAACTAGAGGATGCTCTCTGCTCTTAAAGAAAGATTCCCAACCAATGG 410
 DB 162 erSerGlnGluAlaGluAlaSerSerCysProArgLysLysAspSerSerPrometv 182

QY 411 TGATATGCCAATTAAGAGGGGCACACAATCTATGATATAGCAATTTCTAGAACACAGAG 470
 DB 182 alileCysGlnLeuLysGlyGlyAlaGlnMetLeuLysLysAspAsnGlyAlaArg 202

QY 471 AACTAAAGACTCCATTTGGTCTCTCAGTATCAGATCAAAATTAATTTCTACAGTCAG 530
 DB 202 luLeuLysAlaLeuHisLeuLeuProGlnTyrAspGlnSerSerPheProGlnSerG 222

QY 531 ATCTCCCTAAACCAATGCTCTCTAGTAGGAGATTTTGGCAGCATCAACAAATTA 590
 DB 222 luLeuProLysPrometThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuA 242

QY 591 ATCTCATTAACAACAACCTTGGAGGCGCTTACCATCGGTAGTCAACGGGTCTGCTTCC 650
 DB 242 snLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPhe 262

QY 651 CCTCGGATCACTCTTCAGGACCAACCAAAATTAATTTGGTCTGCTGCTGCTGCTGCT 710
 DB 262 roSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAspCysp 282

QY 711 TTCCAGTGGGACTTTTCCCAACTGCAATTTGTAATTTGTGCAACACTGCATC 770
 DB 282 heSerSerGlyAspPheCysAsnSerCysSer-----CysAsnLeuArgH 298

QY 771 ATCATATTGAACGGTTTAAAGCCATTAAAGCATGCTTGGTGAATTCAGAAAGCTTCC 830
 DB 298 isGluLeuGluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheG 318

QY 831 AGCCAAAAATTGGGAGGCCAATTGGGCAATGTCAAGCCCGCAGCAACAAAGGGTGCA 890
 DB 318 InProLysMetGlyLysGlyValArgLeuGlyAlaAlaLysLeuArgLysSerLysGlyCysA 338

QY 891 ACTGCAGAGGTCAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 950
 DB 338 snCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIleMetC 358

QY 951 GTTCTTCTATTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1010
 DB 358 ysSerSerIleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArgLysM 378

QY 1011 CACTAATGAGCATGCCAAACTACATGCAGCTGAGGCTTGGAGGAGGAGGAGGAGGAGG 1070
 DB 378 etLeuMetSerThrProHisTyrMetGluProGlyAspPheGluSerSerHisTyrLeuS 398

QY 1071 CACCAACGAAATTTCCAGGACTTCCAAAGACTTCAGTCAGTACAGGCGGCTCTCTCT 1130
 DB 398 erProAlaLysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSerCysI 418

QY 1131 TCTCTGGGAGGTGGTGGAGCCACATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1190
 DB 418 leSerTrpGluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluAlaG 438

QY 1191 AGAAACACACTGCTCCTCAAGTGCCTGGCAGACAGATGATCTCTGAGGAAATTTGGAAG 1250
 DB 438 luGlnGluHisCysSerProSerLeuAlaGluGlnMetIleuGluGluPheGlyArgC 458

QY 1251 GCTTATCAGACTTCTCCACTGAGTTTAAATCTTAAGGATTTGAAATTTGAG 1303
 DB 458 ysSerGlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 475

RESULT 2
 Q8VIE0 ID Q8VIE0 PRELIMINARY; PRT; 475 AA.
 AC Q8VIE0;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Tesmin.
 GN TESM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN SEQUENCE FROM N.A.
 RP STRAIN=129;
 RC "Structure of the tesmin gene encoding a testis-specific persistent protein; a possible multifunctional protein with dynamic changes of RT localization throughout spermatogenesis.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB057423; BAB64934.1; -;
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 1.
 SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Alignment Scores:
 Pred. No.: 2,09e-138 Length: 475
 Score: 1508.50 Matches: 289
 Percent Similarity: 78.04% Conservatives: 38
 Best Local Similarity: 68.97% Mismatches: 71
 Query Match: 39.45% Indels: 21
 DB: 11 Gaps: 3

US-09-743-237-3 (1-2134) x Q8VIE1 (1-475)

QY 51 CTCCTCGGGAGTACCCCGGATCCAGAGCTACGCGCTGGAGAGCTGCGCTCTCTG 110
 DB 78 LeuLeuGlyAspPheProGlyLeuProGlyLeuArgSerProAspAlaAala----- 95

QY 111 CAGGCCCGCAGCCCGCCCTGCAACGTGCACCTCTCTCTCTCTCTCTCTCTCTCT 170
 DB 96 -----ProProAlaTyrSerValHisValLeuSerSerLeuLeuProGlyAla 112

QY 171 CGCAGCCCGCGGTGTTTGGCCCTGGGCGCTGGGCTCTGGGAGAGCTCCACCC 230
 DB 113 ArgGlyProAlaLeuLeu-----ProLeu-SerAl 122

QY 231 GGGCGTCCGATGATCCCGAGTTGAATCAAGTAAGCTAGAGTGGTACTACTACAAGTAATA 290
 DB 122 aGlyValArgValIleProValGluIleLysGlu-AlaGlyGlySerValProGlyGlys 142

QY 291 ATCCGGAAGAACCACTTTCGAGATCTTCTGCTCAGCAATCTCTGTCAGTCCCTCCAT 350
 DB 142 erProGlyAspAlaAlaPheGlnAlaProLeuAlaGlnLysSerCysLysPheProS 162

QY 351 GGTCCAGCAACTAGAGGATGCTCTCTGCTCTTAAAGAAAGATTCCCAACCAATGG 410
 DB 162 erSerGlnGluAlaGluAlaSerSerCysProArgLysLysAspSerSerPrometv 182

QY 411 TGATATGCCAATTAAGAGGGGCACACAATCTATGATATAGCAATTTCTAGAACACAGAG 470
 DB 182 alileCysGlnLeuLysGlyGlyAlaGlnMetLeuLysLysAspAsnGlyAlaArg 202

QY 471 AACTAAAGACTCCATTTGGTCTCTCAGTATCAGATCAAAATTAATTTCTACAGTCAG 530
 DB 202 luLeuLysAlaLeuHisLeuLeuProGlnTyrAspGlnSerSerPheProGlnSerG 222

QY 531 ATCTCCCTAAACCAATGCTCTCTAGTAGGAGATTTTGGCAGCATCAACAAATTA 590
 DB 222 luLeuProLysPrometThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuA 242

QY 591 ATCTCATTAACAACAACCTTGGAGGCGCTTACCATCGGTAGTCAACGGGTCTGCTTCC 650
 DB 242 snLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPhe 262

QY 651 CCTCGGATCACTCTTCAGGACCAACCAAAATTAATTTGGTCTGCTGCTGCTGCTGCT 710
 DB 262 roSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAspCysp 282

QY 711 TTCCAGTGGGACTTTTCCCAACTGCAATTTGTAATTTGTGCAACACTGCATC 770
 DB 282 heSerSerGlyAspPheCysAsnSerCysSer-----CysAsnLeuArgH 298

QY 771 ATCATATTGAACGGTTTAAAGCCATTAAAGCATGCTTGGTGAATTCAGAAAGCTTCC 830
 DB 298 isGluLeuGluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheG 318

QY 831 AGCCAAAAATTGGGAGGCCAATTGGGCAATGTCAAGCCCGCAGCAACAAAGGGTGCA 890
 DB 318 InProLysMet

Alignment Scores:		
Prod. NO.:	7.65e-53	Length: 277
Score:	633.50	Matches: 133


```

Qy 199 -----GCGCCTGGGTC 210
Db GlySerIleValProLeuProSerValGlnThrLeuValSerLysArgAlaLeuGlyAla 496
Qy 211 TCGAAGAGGCTCCACCGGGCGTCGATGATCC----- 247
Db 111111 1111111111 111
Qy 497 IleSerAsnAlaSerLysProAlaSerAlaAlaSerSerAlaThrProSerAlaLe 516
Qy 248 -----CAGTTGAATCAAGGTA 264
Db 1111111111
Qy 516 rGlnGluLeuProArgLysHisArgLeuThrAspLeuAsnValGlnLeuLysGlnSerAl 536
Qy 265 AGCAGGTGGTACTACTACAAGTAATAAT-----CCGGAAGAAGCAACTTTGGCA 312
Db 111111 1111111111 111
Qy 536 aSerValSerSerGluAlaSerSerSerAlaGlyProGluAlaLysLysProAr 556
Qy 313 GAATCTTCTGCTCAGGAATCCTTTCAGATGTCCTCCATGGTCCAGGAACATAGAG----- 367
Db 111111 1111111111 111
Qy 556 gTyrValIleThrMetGlnGlnGlySerGlnLysAlaAlaSerGlnProValGlnLysLe 576
Qy 368 -----GATGCTCTCTGCTG 381
Db 1111111111
Qy 576 uIleAsnArgThrAlaAsnValGlnArgValValSerSerThrSerProSerSerAs 596
Qy 382 TTCTCTTAAGAAGATTCCAAACCAATGGTGATGCAATTTGAAGGGGCGCACACAAAT 441
Db 111111 1111111111 111
Qy 596 nSerThrLysLysIleTyrAsnTyrValGlnProThrGlySerAsnGlyAlaLysTyrMe 616
Qy 442 GCTATGT-----ATGACAAATTTAGAACAGAGAACTAAAGCACTCCATTTGGT 492
Db 111111 1111111111 111
Qy 616 tIleCysAsnSerGlyValProGlnGlnSerSerThrSerAlaMetArg----- 631
Qy 493 TCCCTCAGTATCAAGATCAAAATAATATTACAGCTCAGATGTC-----CCTAAACCAAT 546
Db 111111 1111111111 111
Qy 632 -----ArgGlyTyrThrGlyTyrValGluAsnLysThrArgArgProProI 648
Qy 547 GACTGCTTTAGTAGGAGATT----- 568
Db 1111111111
Qy 648 eSerProGlnGlnHisArgPheLysGlnMetGlyProGlnGlnSerLysHisGlnG 668
Qy 569 -TTGCCAGCATCAACAAATAAATTCATACACACAACTT----- 610
Db 111111 1111111111 111
Qy 668 nLeuGlnAlaGlnAlaLysGlnArgIleArgGlnGlnGlnLeuProThrGluGlnSerTh 688
Qy 611 -----CAGGGAGCCTTACCATCGGTA-----GTCAACGGGTC 642
Db 111111 1111111111 111
Qy 688 rProIleLysValGluProLysLeuProThrLeuProProGlyValLysAlaAsnValPr 708
Qy 643 TGCTTTCCCTCGGATCAACTCTTCAGGACCAACCAAAATAACTTTGGCTGGG----- 697
Db 111111 1111111111 111
Qy 708 oAlaLysProLeuPheGluValLeuLysProProAlaThrAlaAlaAlaGlyAlaVa 728
Qy 697 ----- 697
Db 1111111111
Qy 728 lAspProLeuGlyGlyMetThrSerArgArgLysHisCysAsnCysSerLysSerGlnCy 748
Qy 698 -----TACTGTGACTGCTTTGCGAGTGGGACTTTTGCAACAACTGCAATTTGTA 747
Db 111111 1111111111 111
Qy 748 sLeuLysLeuTyrCysAspCysPheAlaAsnGlyLupheCysGlnAspCysThrCysLy 768
Qy 748 TAATTTGTCACAACTTGATGATGATATGACGGTTTAAAGCCATTAAGGCATGTCT 807
Db 111111 1111111111 111
Qy 768 sAspCysPheAsnAsnLeuAspTyrGluValGluArgGluAlaIleArgSerCysLe 788
Qy 808 TGGTAGAATCCAGAACTTCCAGCAAAATTTGGGAAGGGCCAAATGGCAATGTCAA 867
Db 111111 1111111111 111
Qy 788 uAspArgAsnProSerAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetAr 808
Qy 868 GCCCAGCACAACAAAGGTGCAACTGAGGAGGTGAGGTGCTGAGGAATTAAGTGA 927
Db 111111 1111111111 111
Qy 808 g----LeuHisAsnLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGl 827

```

```

Qy 928 GTGCTATGAGGCCCAAAATATTGTGTTCTTCTTATTGCAAAATGCAATGGTGCACAAATTA 987
Db 827 uCysTyrGluAlaLysIleProCysSerSerIleCysLysCysValGlyCysArgAsnMe 847
Qy 988 TGAAGAAGCCCAAGAACGACACTAATGAGCATGCCAACTACATGACAGACTGGAGG 1047
Db 847 tGluAspArgProAsp-----ValAspMetAspSerLeuAspGlyLeuMetGl 863
Qy 1048 TTTGGAAGCAGCCATACCTGCCCAACGAAATTTTCAGGACTTCCAAGATTCACTCA 1107
Db 863 yValGluGlyGln-----LysLysAspLysAlaLysAsnLysGlnLeuAs 878
Qy 1108 CGATAGCGCGCTTCTCATGCTCTCTGGAGGTGGAGGCCACATGCGCTGCCT 1167
Db 878 nGluAsnArgAlaAsnIleTyrPheThrAspValIleGluAlaThrIleMetCysMe 898
Qy 1168 GCTTGCTCAGGAGAGCGCCGAGAAAGACACACTGCTCCAAGTGCCTGGCAGACAGAT 1227
Db 898 tIleSerArgIleValMetHisGluLysGlnAsnValAlaValGluAspMetGluArgGl 918
Qy 1228 GATCCTGGAGGAATTTGGAAGGTCTTATCACAGATTCTC 1267
Db 918 uValMetGluGluMetGlyGluSerLeuThrGlnIleIle 931

RESULT 7
Q95QD7 PRELIMINARY; PRT; 435 AA.
AC Q95QD7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Jc8.6a protein.
GN Jc8.6 OR Jc8.6A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lightning J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z82274; CAB05229.1; -.
DR WormPep; Jc8.6a; CEI7989.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
SQ SEQUENCE 435 AA; 49109 MW; 7629A5C749D25A5D CRC64;

Alignment Scores:
Pred. No.: 2,73e-34 Length: 435
Score: 444.50 Matches: 130
Percent Similarity: 43.00% Conservative: 48
Best Local Similarity: 31.40% Mismatches: 116
Query Match: 11.62% Indels: 120
DB: 5 Gaps: 19

US-09-743-237-3 (1-2134) x Q95QD7 (1-435)
Qy 142 ACTTCCTGCTCTGCTGCTAC-----CCGCGCAGCGAGCCCGGGGTGTTTCCC 192
Db 78 ThrSerHisArgValTyrValProProArgProValGlnArgLysLeuPheGln 97
Qy 193 CTTGGG-----GGCCTGGTCTCGGAGGAGGACCTCCACCGGGCGTCCGATGATC 246
Db 98 ProGlyProSerThrProGlySerSerGln----- 107

```

```
QY 247 CCAGTTGAAATCAAGGTAAGCAGGTGGTACTACTACAGTAATAATCAATCCGGAAGAACAC 306
Db 108 -----Tyr-ThrValArgAsn----- 112
QY 307 TTTCAGAAATCTCTTCGTCAGGAATCTGTTGCAAGTTCCCATGGTCCAGGAAGACTAGA 366
Db 113 -LeuSerAsnLeuSerGly-----SerProSerMetTyr 123
QY 367 GGATGCTCCTGCTGCTCTCTTAAGAAAGATCCCAACCAATGCGTGTATATGCAATTTGAA 426
Db 123 rAspArgGlnProAlaSerLeuProArgThrValGlnProMet----- 137
QY 427 AGGGGACACAAATGCTATGATAGACAATTTTAGACAAGAAAGAACTAAAGCACTCCA 486
Db 138 ---GlyLeuGluMet-----GlyAsnSerGluGlnArgLysValTyrIle----- 151
QY 487 TTGGTTCCTCAGTATCAAGATCAAAATAATATCTACAGTCAGATGTCCTTAACCAAT 546
Db 152 -----AspMetLysAspHisValSerHisIleArgLeuLysThrLysLysVal 168
QY 547 GACTGCTTTAGTAGGAGATTTTGGCAGCAGCATCAACAAAATTAATCTCATACACAACA 606
Db 168 l-----PheAlaProGlyGlnArgLys----- 175
QY 607 ACTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTTCCCTCGGATCAACTCT 666
Db 176 -----ProCysAsnCysThrLys 181
QY 667 TCAGGACACCAAAATAACTTTGCTGGTACTGTGCTGCTTTGGCAGTGGGACTT 726
Db 181 sSerGlnCysLeuLysLeu-----TyrCysAspCysPheAlaAsnGlyGluPh 197
QY 727 TTGCAACAACTGCAATGTAATATGTTGCAACAACTTGCATCATGATATCAACGGTT 786
Db 197 eCysArgAspCysAsnCysLysAspCysHisAsnAsnIleGluTyrAspSerGlnArgSe 217
QY 787 TAAAGCCATTAAGGATGCTCTGGTGAAGAAATCAGAAAGCTTCCAGCCAAAATTTGG-- 844
Db 217 rLysAlaIleArgGlnSerLeuGluArgAsnProAsnAlaPheLysProLysIleGlyl 237
QY 845 ----AAGGCCAATTTGGCAATGTCAAGCCCGAGCAACAAGGTGCAACTGCAGGAG 900
Db 237 eAlaArgGlyIleThrAspIleGluArgLeuHisGlnLysGlyCysHisCysLysLys 257
QY 901 GTCAGGCTCCCTGAAGAAATTAATGCGAGTGTATGAGGCCCAAAATTAATGTTCTCTAT 960
Db 257 sSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysValProCysThrAsp 277
QY 961 TTGCAAAATGCTGTTGCAAAAT-----TATGAAGAAAG 996
Db 277 gCysLysCysLysGlyCysGlnAsnThrGluThrTyrArgMetThrArgTyrLysAsnSe 297
QY 997 C-----CCAGACGAAGACACTAATGAGCATGCCAAACTACATGCAGACT-- 1042
Db 297 rGlyGlyAlaValSerAsnThrAsnAlaLeuMetSerLeuThrAsnAlaSerSerThrAl 317
QY 1043 -----GGAGGTTTG-----GAAGGCGCCATTACCTGCC 1071
Db 317 aThrProAspSerGlyProGlySerValValThrAspGluHisGlyAspTyr---Gl 336
QY 1072 ACCAAGAAATTTTCAGGACTTCCAGAGATTCAGTCACGATAGCGCG-----CCTTCCTC 1125
Db 336 uAspMetLeuLeuSerHisLysProLysValGluMetAspProArgPheProTyr 356
QY 1126 ATGCATCTCTGGAGGTGGTGGCCACATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1185
Db 356 rTyrMetThrAspGluValGluAlaAlaThrMetCysMetValAlaGlnAlaGlu 376
QY 1186 GGCC-----GAGAAAGAACTGCTCCAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1227
Db 376 uAlaLeuAsnTyrGluLysValGlnThrGluAspGluLysLeuIleAsnMetGluLysLe 996
```

```
QY 1228 GATCCTGGAGGAATTTTGAAGTGTCTTATCAGATTTCTC 1267
Db 396 uValLeuArgGluPheGlyArgCysLeuGluGlnMetIle 409

RESULT 8
O62295 PRELIMINARY; PRT; 429 AA.
ID AC O62295;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Jc8.6b protein.
GN Jc8.6 OR Jc8.6B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lightning J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA none;
RX MEDLINE=99069613; PubMed=9851916;
RT "Genome sequence of the nematode C. elegans: A platform for
RL Science 282:2012-2018(1998).
DR EMBL; 282274; CAB05228.1; -.
DR WormPep; Jc8.6b; CEI1990.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 429 AA; 48365 MW; 10601B22681C06D2 CRC64;

Alignment Scores:
Pred. No.: 1.47e-33 Length: 429
Score: 437.00 Matches: 92
Percent Similarity: 56.95% Conservative: 35
Best Local Similarity: 41.26% Mismatches: 62
Query Match: 11.43% Indels: 34
Db: 5 Gaps: 8

US-09-743-237-3 (1-2134) x O62295 (1-429)
QY 698 TACTGTGACTGCTTCCAGTGGGACTTTTGCAACAACATGCAATTTGTAATTTGTTGC 757
Db 182 TyrCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHis 201
QY 758 AACAACTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGTCTGTGTAGAAAT 817
Db 202 AsnAsnIleGluTyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 221
QY 818 CCAGAAGCTTTCAGCCAAATTTGGG-----AAGGGCAATTTGGCAATGTCAGGCC 871
Db 222 ProAsnAlaPheLysProLysIleGlyIleAlaArgGlyIleThrAspIleGluArg 241
QY 872 CAGCAACAAGGGTGCAGTGCAGGCTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 931
Db 242 LeuHisGlnLysCysHisCysLysSerGlyCysLeuLysAsnTyrCysGluCys 261
QY 932 TATGAGGCCAAATTAATGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 985
Db 262 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 281
QY 986 -----TATGAAGAAAG-----CCAGAACGAAGAACACTA 1015
Db 282 ThrTyrArgMetThrArgTyrLysAsnSerGlyAlaValSerAsnThrAsnAlaLeu 301
QY 1016 ATGAGCATGCCAACTACATGCAGACT-----GGAGGTTTG--- 1051
Db 302 MetSerLeuThrAsnAlaSerSerThrAlaThrProAspSerGlyProGlySerVal 321
QY 1052 -----GAAGGCGCCATTACCTGCCAACCAAGAAATTTTTCAGGACTTCCAGATTC 1102
```


OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Magnoliaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao Y., Ishmael N., Kumar N., Redman J., Smith S., Riedmuller S.,
RT "Cloning and sequencing of full-length cDNAs for hypothetical genes
RT from chromosome 2 of Arabidopsis thaliana."
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY163775; AA85199.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 532 AA; 57033 MW; 78CF4E441175958D CRC64;

Alignment Scores:
Pred. No.: 3 59e-27 Length: 532
Score: 372.00 Matches: 94
Percent Similarity: 44.51% Conservative: 48
Best Local Similarity: 29.47% Mismatches: 91
Query Match: 9.73% Indels: 86
DB: 10 Gaps: 10

US-09-743-237-3 (1-2134) x Q8GUZ0 (1-532)

QY 527 TCAGATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTTCAGCATCAACAAA 586
Db 16 AlaserileProSerProileValThrValThrArgProileThrSerGlnAlaPro 35
QY 587 TTAATCTCATTAACA-----CAACAACCTGAGGAGCCTTA----- 622
Db 36 ProThrValAlaThrProileProProProGlnSerGlnGlyLeuHisVal 55
QY 623 -----CCATCGTAGTCACAGGCTGCTTTCCCTCGGATCAACT 664
Db 56 ProileArgHisProArgProGluSerProAsnSerMetProArgProAlaGlyGluThr 75
QY 665 CTTCCAGGACCAACAAA-----ATACTTTGGCT 694
Db 76 ArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHisSerArgCysLeuLys 95
QY 695 GGGTACTGTGACTGCTTCCAGTGGGACTTTTCACACACTGCAATTTGTAATGCT 754
Db 96 LeuTyrcysGluCysPheAlaSerGlyThrTyrcysAspGlyCysAsnCysValAsnCys 115
QY 755 TGCAACAACCTGCATCATGATATTCAGCGTTTAAAGCCATTAAAGCATGCTCTGGTAGA 814
Db 116 PheAsnAsnValGluAsnGluProAlaArgArgGlnAlaValGluSerThrLeuGluArg 135
QY 815 AATCCAGAGCTTCCAGCCAAAAT-----GGGAG 847
Db 136 AsnProAsnAlaPheArgProLysileAlaAlaSerProHisGlyArgAspAsnArg 155
QY 848 GCCCAATTTGGCAATGTC-----AAGCCAGCAGCACAAGAGGTGCACAGGAGG 901
Db 156 GluGluValGlyAspValMetLeuAlaArgHisAsnLysGlyCysHisCysLysLys 175
QY 902 TCAGGCTGCTCAAGAAATTTACTGCGAGTGCTATGAGCCCAATATGTTCTCTATT 961
Db 176 SerGlyCysLeuLysTyrcysGluCysPheGlnAlaAsnIleLeuCysSerGluAsn 195
QY 962 TGCAATGCTGTTGGTGAATAATTTAAGAAAGCCAGAACACATTAATGAGC 1021
Db 196 Cys**CysLeuAspCysLysAsnPheGluGlySerGluValArgGlnSerLeuPheHis 215
QY 1022 ATGCCAAC-----TACATGCAG-----ACTGGAGT 1048
Db 216 GlyGluHisSerHisAsnLeuAlaTyrlLeuGlnHisAlaAsnAlaAlaIleThrGlyAla 235
QY 1049 TTGGAGGAGCCATTACCTG-----CCACCAACCAATTTTCAGGA----- 1090
Db 236 IleGlySerSerGlyPheAlaSerAlaProProLysArgArgLysGlyGlnGluIle 255

QY 1091 -----CTTCAAGATTTCAGTCACGAT 1111
Db 256 PhePheAsnGlnGlyThrLysAspSerSerThrHisArgLeuGlyGlnAlaAsnAsnGly 275
QY 1112 AGCGCGCTTCTCTCA----- 1126
Db 276 ArgThrThrSerSerGlnThrGlySerArgAlaGlyGlyAsnAlaSerLeuGlyProSer 295
QY 1127 -----TGCATCTCCTGGGAGGTGGTGAGGCCACATGCGCCCTGCCG 1168
Db 296 LysSerLeuLeuAlaAsnIleIleLysProMetAspValLysAlaLeuCysSerValLeu 315
QY 1169 CTTGCTCAGGAGAGAGCGGAGAACACTGCTCCAGCTGCTGCGAGAGCAG 1225
Db 316 ValAlaValAlaGlyGluAlaAlaLysThrLeuThrGluLysArgLeuAlaAsnGln 334

RESULT 11
Q8GUZ2 PRELIMINARY; PRT; 536 AA.
ID AC Q8GUZ2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN T2G17.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Magnoliaceae; Arabidopsis.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiao Y., Ishmael N., Kumar N., Redman J., Smith S., Riedmuller S.,
RT "Cloning and sequencing of full-length cDNAs for hypothetical genes
RT from chromosome 2 of Arabidopsis thaliana."
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY163773; AA85197.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 536 AA; 57522 MW; 3A841C7222010E08 CRC64;

Alignment Scores:
Pred. No.: 5 65e-27 Length: 536
Score: 370.00 Matches: 94
Percent Similarity: 43.96% Conservative: 48
Best Local Similarity: 29.10% Mismatches: 91
Query Match: 9.68% Indels: 90
DB: 10 Gaps: 10

US-09-743-237-3 (1-2134) x Q8GUZ2 (1-536)

QY 527 TCAGATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTTCAGCATCAACAAA 586
Db 16 AlaserileProSerProileValThrValThrArgProileThrSerGlnAlaPro 35
QY 587 TTAATCTCATTAACA-----CAACAACCTGAGGAGCCTTA----- 622
Db 36 ProThrValAlaThrProileProProProGlnSerGlnGlyLeuHisVal 55
QY 623 -----CCATCGTAGTCACAGGCTGCTTTCCCTCGGATCAACT 664
Db 56 ProileArgHisProArgProGluSerProAsnSerMetProArgProAlaGlyGluThr 75
QY 665 CTTCCAGGACCAACAAA-----ATACTTTGGCT 694
Db 76 ArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHisSerArgCysLeuLys 95
QY 695 GGGTACTGTGACTGCTTTCAGGAGGACTTTTCACCAACTGTAATTTGTAATGCT 754
Db 96 LeuTyrcysGluCysPheAlaSerGlyThrTyrcysAspGlyCysAsnCysValAsnCys 115
QY 755 TGCAACAACCTGCATCATGATATTCAGCGTTTAAAGCCATTAAAGCATGCTCTGGTAGA 814
Db 116 PheAsnAsnValGluAsnGluProAlaArgArgGlnAlaValGluSerThrLeuGluArg 135
QY 815 AATCCAGAGCTTCCAGCCAAAAT-----GGGAG 847
Db 136 AsnProAsnAlaPheArgProLysileAlaAlaSerProHisGlyArgAspAsnArg 155
QY 848 GCCCAATTTGGCAATGTC-----AAGCCAGCAGCACAAGAGGTGCACAGGAGG 901
Db 156 GluGluValGlyAspValMetLeuAlaArgHisAsnLysGlyCysHisCysLysLys 175
QY 902 TCAGGCTGCTCAAGAAATTTACTGCGAGTGCTATGAGCCCAATATGTTCTCTATT 961
Db 176 SerGlyCysLeuLysTyrcysGluCysPheGlnAlaAsnIleLeuCysSerGluAsn 195
QY 962 TGCAATGCTGTTGGTGAATAATTTAAGAAAGCCAGAACACATTAATGAGC 1021
Db 196 Cys**CysLeuAspCysLysAsnPheGluGlySerGluValArgGlnSerLeuPheHis 215
QY 1022 ATGCCAAC-----TACATGCAG-----ACTGGAGT 1048
Db 216 GlyGluHisSerHisAsnLeuAlaTyrlLeuGlnHisAlaAsnAlaAlaIleThrGlyAla 235
QY 1049 TTGGAGGAGCCATTACCTG-----CCACCAACCAATTTTCAGGA----- 1090
Db 236 IleGlySerSerGlyPheAlaSerAlaProProLysArgArgLysGlyGlnGluIle 255


```

QY 1217 GCAGAGCAG 1225
Db 336 AlaAsnGln 338

RESULT 13
Q9SZD1
AC Q9SZD1 PRELIMINARY; PRT; 603 AA.
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative transcription factor.
GN F19B15.30 OR AT4G29000 OR AT4G29000/F19B15_30.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL078470; CAB43914.1; -
DR EMBL; AL161574; CAB79658.1; -
DR EMBL; AK118658; BAC43254.1; -
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 603 AA; 64635 MW; 7FD1951AF818DE5 CRC64;

Alignment Scores:
Pred. No.: 8,748-26 Length: 603
Score: 358.00 Matches: 93
Percent Similarity: 44.52% Conservative: 41
Best Local Similarity: 30.90% Mismatches: 76
Query Match: 9.36% Indels: 91
DB: 10 Gaps: 12

US-09-743-237-3 (1-2134) x Q9SZD1 (1-603)

QY 521 CTACAGTCAGATGTCCT-----AAACCA-----ATGACTGCTTTAGTAGGGAGATTT 568
Db 58 IleSerSerValProSerThrIleArgProGlyMetThrIleAlaIleGlyGlnVal 77
QY 569 -----TTGCCAGCATCAACAAATAATCTCATTCACACAACTT 610
Db 78 ThrGlnValArgProThrLeuProMetAlaThrThrMetSerAsnProProSerGlnSer 97
QY 611 GAGGAGCCTTACCATCGTAGTCAACGGGTCTGCTTCCCTCGGATCACTCTTCCA 670
Db 98 Gln-----IleValAsnAla-----ProIleArgHisProIlePro 109

671 GGACCACCAAAATA----- 685
110 GluSerProLysAlaArgGlyProArgProAsnValGluGlyArgAspGlyThrProGln 129
686 -----ACTTTGGTGGGTACTGTGACTGCTTT 712
130 LysLysLysGlnCysAsnCysLysHisSerArgCysLeuLysLeuTyrCysGluCysPhe 149
713 GCCAGTGGGACTTTTGAACAACTGCAATTGTAATATTGTAACAACTGTCATCAT 772
150 AlaSerGlyThrTyrCysAspGlyCysAsnCysValAsnCysPheAsnValAspAsn 169
773 GATATTGAACGGTTTAAAGCCATTAGGCATCTCTTGGTAGAAATCCAGAACTTTCCAG 832
170 GluProAlaArgArgGluAlaValGluAlaThrLeuGluArgAsnProPheAlaPheArg 189
833 CCAAAATTTGGGAAGGCCAATTGGGCAAT----- 862
190 ProLysIleAlaSerSerProHisGlyArgAspLysArgGluAspIleGlyGluVal 209
863 ---GTCAAGCCCCCAGCACAAAGGGTGCACACTGCAGGAGTCAAGGCTGCTGAGAAAT 919
210 ValLeuLeuGlyLysHisAsnLysGlyCysHisCysLysLysSerGlyCysLeuLysLys 229
920 TACTCGAGTGTCTATGAGGCCCAATATGCTTCTTCTTCTTATTTGCAAAATGCAATTGCTGC 979
230 TyrCysGluCysPheGlnAlaAsnIleLeuCysSerGluAsnCysLysCysLeuAspCys 249
980 AAAAATTATGAAGAAAGCCAGACGAAAGACACTAATAGCATGCCAAAC----- 1030
250 LysAsnPheGluGlySerGluGluArgGlnAlaLeuPheHisGlyGluHisSerAsnHis 269
1031 -----TACATGCCAG-----ACTGGAGGTTTGGAAAGCCACCAT 1063
270 MetAlaTyrLeuGlnGlnAlaAlaAsnAlaAlaIleThrGlyAlaValGlySerSerGly 289
1064 TACCTGCCA-----CCAACGAAATTTTCAGGA----- 1090
290 PheAlaProSerProAlaProLysArgLysGlyGlnGluLeuPheAsnGlnAla 309
1091 CTTCCAGATTCAGTCAGATAGCGGCTTCTCTCATGTCATCTCTGGAGGTGGAG 1150
310 Ile-LysAspSerSerArgLeuSerHisPheProGlnValAsnAsnGly----- 325
1151 GCCACATCGCTGCTGCTTCTCAGGAGGAGAGCGCGAGAAAGAAACACTGCTCCCAAG 1210
326 -----ArgThrGlyGlyProThrSerGlyThrSerProse 337
1211 T 1211
337 r 337

RESULT 14
Q9CAV1
ID Q9CAV1 PRELIMINARY; PRT; 601 AA.
AC Q9CAV1;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical 66.8 kDa protein.
GN T9J14.20.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,

```

```

RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choise N., Artiguenave F., Robert C., Bottier P.,
RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Schaefer M., Schoen O., Barques M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA De Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shimpō S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:820-822(2000).
DR EMBL: AC009465; AAG51411.1; -.
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
KW Hypothetical protein.
SQ SEQUENCE 601 AA; 66786 MW; 10848E970D81E022 CRC64;

Alignment Scores:
Pred. No.: 1,23e-16 Length: 601
Score: 264.50 Matches: 81
Percent Similarity: 36.95% Conservative: 28
Best Local Similarity: 27.46% Mismatches: 101
Query Match: 6.92% Indels: 85
DB: 10 Gaps: 12

US-09-743-237-3 (1-2134) x Q9CAV1 (1-601)
QY 314 AATCTCTGTGCTAGGAATCTGTGCAAGTCCCATGTCCTCCAGCACTAGAGGATGCC 373
Db 255 AsnLeuSerCysSerSerLysValAlaAlaIleAspSerThrAlaGluAlaGluAspLys 274
QY 374 TCCTGCTCTTCTTGAAGAAGATCCNACCAATGGTGATATGCCAATTCGAAAGGGGCC 433
Db 275 Glu-----AspLysAspLeuGlnPro-----SerGly 283
QY 434 ACACAAATGCTATGTATAGCAATCTAGAACAGAAAGCACTCCATTTGGTT 493
Db 284 LysGlnArgSerValArgArgCysLeuThrPheAspMetGlyGlySerHisLysArg 303
QY 494 CCTCAGTATCAAGTCAAAATAATATCTACAGTCAGAT-----GTC 535
Db 304 IleProLeuArgAspSerThrAsnAspLeuProLeuAspSerThrSerIleAsnLysAla 323
QY 536 CCTAACCAATGACTGCTTTA-----GTAGGAGATTTTCCGACCA 577
Db 324 ProSerProGlnAsnCysLeuAspThrSerLysGlnAspThrAspGluLeuLeuProIle 343
QY 578 TCAACAAATTAATCTATTACACAACTCTGAGGAGCGCTTA---CCATCGGTA--- 631
Db 344 ProArgThrIleGlyLeu-----HisLeuAsnGlyPheValAsnProSerValSer 360
QY 632 -----GTCAACGGGTGCTTTCCCTCGGGATCA----- 661
Db 361 SerGlyArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 380
QY 661 ----- 661

```

```

Db 381 TyrAsnIleGluAspGluPheSerThrProValSerThrLysArgAspLeuValPhe 400
QY 662 -----ACTCTTCCAGGACACCAACCAATACTTTGGCTGGG----- 697
Db 401 SerAspValIleMetGluProGluArgSerValGluGlyGluCysPheAspGln 420
QY 698 -----TACTGTGACTGCTTTGCCAGTGGGAGACTTTTGCAACAAC 736
Db 421 LeuMetaIleMetGluAsnArgTy-CysGluCysPheSerAlaGlyLeuPheCysGlyGlu 440
QY 737 ---TGCATTTCTAATAATTGTTGCAACAACCTTGCATCATGATATATGAACGGTTAAAGCC 793
Db 441 ProCysSerCysGlnAsnCysPheAsnLysProIleHisGluAspLeuValMetLysSer 460
QY 794 ATTAAGGCATGCTTGTGTAGAAATCCAGAACCTTTCCAGCCCAAAATTTGGGAAGGCCAA 853
Db 461 ArgGluValIleLysAlaArgAsnProLeuAlaPheAlaProLysValValSerThrSer 480
QY 854 -----TTGGGCAATGCTCAAG-----CCCCAGCACCAAC 880
Db 481 AspThrValIleAspLeuTirpValGluAsnSerLysThrProAlaSerAlaArgHisLys 500
QY 881 AAAGGGTCCAACTCCAGGAGCTCAGGCTGCTGAGAAATTTACTCGGAGTGTATGAGGCC 940
Db 501 ArgGlyCysAsnCysArgLysSerGlyCysSerLysLysTyrcysGluCysPheMetMet 520
QY 941 CAAATTATGTCTTCTCTATTGCAAAATGCAATGCTTGGTTGCAAAAT 985
Db 521 GlyValGlyCysSerSerAsnCysArgCysMetGlyCysLysAsn 535

RESULT 15
Q9LW71
ID Q9LW71 PRELIMINARY; PRT; 356 AA.
AC Q9LW71.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genomic DNA, chromosome 3, pl clone: MSU1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucrotis II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty pl and TAC
RL DNA Res. 7:131-135(2000).
DR EMBL; AB012247; BAB02682.1; -.
DR InterPro: IPR005172; CXC.
DR InterPro: IPR000345; CytC_heme_bind.
DR Pfam; PF03638; CXC; 2.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
SQ SEQUENCE 356 AA; 39796 MW; 4FE0EA2D9DBFF58 CRC64;

Alignment Scores:
Pred. No.: 1.67e-16 Length: 356
Score: 262.50 Matches: 56
Percent Similarity: 45.34% Conservative: 17
Best Local Similarity: 34.78% Mismatches: 47
Query Match: 6.86% Indels: 41
DB: 10 Gaps: 5

US-09-743-237-3 (1-2134) x Q9LW71 (1-356)

```

```
QY 698 TACTGTGACTGCTTTCCAGTGGGACTTTTCACACAACTGCAATGTGTAATATTGTC 757
Db |||||||||||||||||||
QY 79 TyrCysAspCysPheAlaSerGlyValValCysThrAspCysAspCysValAspCysHis 98
Db ||| :|||:||||| :|||
QY 758 ACACACTTCATCATGATATTGAACGGTTTAAGCCATTAAAGCATGTCTTGGTAGAAAT 817
Db ||||| ||| :|||: ||||| |||||
QY 99 AsnAsnSerGluLysCysAspAlaArgGluAlaAlaMetValAsnValLeuGlyArgAsn 118
Db ||||| ||| :|||: ||||| |||||
QY 818 CCAGAGCTTTCCAGCCAAATTTGGGAAGGCCCAATTGGCAATGTCAGCCCCCAGCAC 877
Db ||| ||||| ||| :|||: ||||| :|||
QY 119 ProAsnAlaPheSerGluLys-----AlaLeuGlySerLeuThrAspAsnGln 134
QY 878 -----ACAAAGGTTGCAACTGCAGGAGG 901
Db :|||: ||||| ||||| :|||
QY 135 CysLysAlaAlaProAspThrLysProGlyLeuLeuSerArgGlyCysLysCysLysArg 154
Db :|||: ||||| ||||| :|||
QY 902 TCAGGCTGCTGAAGAAATTACTGCGAGTGCTATGAGGCCCAATATATGTGTTCTTCTATT 961
Db :||| :||| ||||| :|||: ||||| :|||
QY 155 ThrArgCysLeuLysLysTyrCysGluCysPheGlnAlaAsnLeuLeuCysSerAspAsn 174
Db :||| :||| ||||| :|||: ||||| :|||
QY 962 TCAAAATGCTGGTTGCCAAAATTTATGAAGAACCCAGAACGAAGACACTAATGAGC 1021
Db ||||||||| ||||| ||||| :|||
QY 175 CysLysCysIleAsnCysLysAsnValSerGluAla----- 186
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTTGGGAAGCGAGCATTTACCTGCCACCAAGGAAA 1081
Db :||| :|||
QY 187 -----PheGlnProProAla--- 191
QY 1082 TTTTCA-----GGACTTCCAAGATTCACGTACGATAGCGGCGCTTCCTCATGCAATCCTCC 1135
Db ||||| :|||: |||||
QY 192 PheSerAlaHisAsnSerProGlnValTyrArgArgArgAspArgGluLeuThrGlu 211
QY 1136 TGG 1138
Db |||
QY 212 Trp 212
```

Search completed: July 24, 2003, 14:30:29
Job time : 212.427 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run On: July 24, 2003, 13:47:53 ; Search time 33.9001 Seconds
(without alignments)
5920.628 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 3824
Sequence: 1 aattcggggtcaagcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/csqn2_1/USPTO_spool/US09743237/runat_24072003_113104_17660/app_query.fasta_1.4750
-DB=SwissProt_41 -FMF=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 @CGN.1.1.111@runat_24072003_113104_17660 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1647	43.1	299	1 MTL5_HUMAN	Q9v415 homo sapien
2	1251	32.7	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
3	130.5	3.4	3695	1 LMA5_HUMAN	O15230 homo sapien
4	123	3.2	1700	1 BAR3_CHITE	Q03376 chironomus
5	119.5	3.1	1172	1 AHM2_ARATH	O64474 arabidopsis
6	119.5	3.1	3010	1 POLG_HCVJA	P26662 h genome po
7	119	3.1	1576	1 YLK3_CAEEL	P41951 caenorhabdi
8	118.5	3.1	1801	1 LMB2_RAT	P15800 rattus norv
9	118.5	3.1	5376	1 ZAN_MOUSE	O88799 mus musculus
10	117	3.1	3011	1 POLG_HCVI	P26664 h genome po
11	117	3.1	3718	1 LMA5_MOUSE	Q61001 mus musculus
12	116.5	3.0	2003	1 NTC4_HUMAN	Q99466 homo sapien
13	116	3.0	3010	1 POLG_HCVBK	P26663 h genome po
14	116	3.0	3010	1 POLG_HCVJT	Q00269 h genome po
15	115.5	3.0	769	1 ITB2_HUMAN	P05107 homo sapien
16	114.5	3.0	778	1 TAST_HUMAN	Q12815 homo sapien
17	114.5	3.0	3726	1 TRX_DROME	P20659 drosophila
18	112.5	2.9	3133	1 HMC2_BOMMO	P98092 bombyx mori

ALIGNMENTS

RESULT 1

ID	MTL5_HUMAN	STANDARD;	PRT;	299 AA.
AC	Q9Y415;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).			
DE	MTL5.			
GN	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=99208669; PubMed=10191092;			
RA	Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;			
RT	"A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation.";			
RL	Genomics 57:130-136(1999).			
CC	!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.			
CC	!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; U86074; AAD24668.1; -			
DR	Genew; HGNC:7446; MTL5.			
DR	MIM; 604374; -			
DR	GO; GO:0005505; F.heavy metal binding activity; TAS.			
DR	GO; GO:0006875; P.metal ion homeostasis; TAS.			
DR	GO; GO:0009412; P.response to heavy metal; TAS.			
DR	GO; GO:0007283; P.spermatogenesis; TAS.			
DR	InterPro; IPR005172; CXC.			
DR	Pfam; PF03638; CXG; 1.			

P29846 h genome po
Q61789 mus musculus
P58660 mus musculus
P28010 homo sapien
Q99066 homo sapien
P19467 mus musculus
Q29131 tupaiia glis
P08169 bos taurus
O88516 mus musculus
Q04592 mus musculus
O88799 mus musculus
P57999 oryctolagus
Q61508 mus musculus
P54643 dictyosteli
P07942 homo sapien
O97776 eulemur ful
P14585 caenorhabdi
O88671 rattus norv
O95980 homo sapien
Q9nyj7 homo sapien
Q9et61 rattus norv
P53714 sus scrofa
P1046 drosophila
O17514 caenorhabdi
P22199 rattus norv
P35448 xenopus lae
P00522 drosophila

19 112 2.9 3010 1 POLG_HCVTW
20 112 2.9 3333 1 LMA3_MOUSE
21 111.5 2.9 1021 1 CARA_MOUSE
22 111 2.9 798 1 ITB7_HUMAN
23 110 2.9 1587 1 LMG3_HUMAN
24 109.5 2.9 573 1 C114_MOUSE
25 109 2.9 977 1 MCR_TUPGB
26 109 2.9 2499 1 MPRI_BOVIN
27 108.5 2.8 592 1 DLL3_MOUSE
28 108.5 2.8 1877 1 PCK5_MOUSE
29 108 2.9 5376 1 ZAN_MOUSE
30 107.5 2.8 2282 1 ZAN_RABIT
31 107 2.8 559 1 ECM1_MOUSE
32 107 2.8 677 1 SP87_DICDI
33 106.5 2.8 1786 1 LMB1_HUMAN
34 106 2.8 884 1 ANDR_EULFC
35 106 2.8 1429 1 L112_CAEEL
36 105.5 2.8 589 1 DLL3_RAT
37 105.5 2.8 971 1 RECK_HUMAN
38 105 2.7 618 1 DLL3_HUMAN
39 105 2.7 643 1 CO93_RAT
40 104.5 2.7 769 1 ITB2_PIG
41 104.5 2.7 1790 1 LMB1_DROME
42 104 2.7 773 1 MES2_CAEEL
43 104 2.7 981 1 MCR_RAT
44 104 2.7 1173 1 TSPL_XENLA
45 104 2.7 1520 1 ABL_DROME

```
KW Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH.
SQ SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;

Alignment Scores:
Pred. No.: 5,68e-126 Length: 299
Score: 1647.00 Matches: 299
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.07% Indels: 0
DB: 1 Gaps: 0

US-09-743-237-3 (1-2134) x MTL5_HUMAN (1-299)
Qy 407 ATGCTGATATGCCAATTTGAAGGGGGGCACAAATGCTATATAGACAATTCAGAAC 466
Db 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
Qy 467 AGAAGCAATAAGCACATCCATTTGGTTCCTCAGTATCAAGATCAAAATAATATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGln 40
Qy 527 TCAGATGCTCCCTAACCAATGACTGCTTAGTAGGGAGATTTTCCAGCATCAACAAA 586
Db 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
Qy 587 TTAATCTCTATTACACAACTTTGAGGGAGCCCTTACCATCGGTAGTCAACGGGCTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
Qy 647 TTCCCTCGGGATCACTTTCCAGGACCCACCAAAATAACTTTGGCTGGGTACTGTGAC 706
Db 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
Qy 707 TGCTTTCCAGTGGGACTTTTGCAACAACCTGCAATTTGTAATTTGTCACAACTTG 766
Db 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnGlySerValValAsnGlySerAla 120
Qy 767 CATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGCTCTTGGTAGAAATCCAGAAGCT 826
Db 121 HisHisAspIleGluAlaGlyPheLysAlaIleLysAlaCysLeuGlyAlaGlnIle 140
Qy 827 TTCAGCCAAAATTTGGAGGGCCAAATTTGGCAATGTCAGCCCGCCAGCACAAAGGG 886
Db 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
Qy 887 TGCAACTGCAGAGGTCAGGTCGCTGAAGAATTTACTGCGAGTCTATGAGGCCCAAT 946
Db 161 CysAsnCysArgArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGlnIle 180
Qy 947 ATGCTTCTTCTATTTCGAAATGCAATGCTGTTGCAAAATATATGAAGAACCCAGACGA 1006
Db 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200
Qy 1007 AAGACATAATGACATGCCAAATACATGACAGATCGAGGTTTGGAGGCGAGCCATTAC 1066
Db 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
Qy 1067 CTGCCACCAACGAATTTTTCAGGACTTCCAAAGATTCAGTACAGTACGAGCCGCTTCCTCA 1126
Db 221 LeuProThrLysPheSerGlyLeuProArgPheSerHisAspArgProSerSer 240
Qy 1127 TGCATCTCTGGGAGGTGGAGGCCACATGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1186
Db 241 CysIleSerTrpGluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGlu 260
Qy 1187 GCCGAGAAGAACACTGCTCCAAAGTGCCTGCCAGAGCAGATGATCCTGGAGGATTTGGA 1246
Db 261 AlaGluLysGluHisCysSerLysCysLeuAlaGlnMetIleLeuGluGluPheGly 280
Qy 1247 AGGTGCTTATCACAGATTTCCACACTGAGTTTAAATCTAAGGGATTGAAATGGAG 1303
Db 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 2
MTL5_MOUSE
ID MTL5_MOUSE STANDARD; PRT; 295 AA.
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RT early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999)
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
CC PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U77383; AAD24667.1; -;
DR EMBL; U67176; AAD24666.1; -;
DR MGD; MGI:1340029; MTL5.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
DR Spermatogenesis.
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Alignment Scores:
Pred. No.: 8,62e-94 Length: 295
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 32.71% Indels: 4
DB: 1 Gaps: 1

US-09-743-237-3 (1-2134) x MTL5_MOUSE (1-295)
Qy 407 ATGCTGATATGCCAATTTGAAGGGGGGCACAAATGCTATATAGACAATTCAGAAC 466
Db 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
Qy 467 AGAAGCAATAAGCACATCCATTTGGTTCCTCAGTATCAAGATCAAAATAATATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrAspAspGlnSerSerPheProGln 40
Qy 527 TCAGATGCTCCCTAACCAATGACTGCTTAGTAGGGAGATTTTCCAGCATCAACAAA 586
Db 41 SerGluLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
Qy 587 TTAATCTCTATTACACAACTTTGAGGGAGCCCTTACCATCGGTAGTCAACGGGCTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
Qy 647 TTCCCTCGGGATCACTTTCCAGGACCCACCAAAATAACTTTGGCTGGGTACTGTGAC 706
```


QY 455 AATTCTAGAACAGAGAACTAAAGCACTCCATTTGGTTCTCTAGTATCAAGATCAAAAT 514
 DB 236 AsnGlyArgProGlyAlaMet----- 242
 QY 515 AATTATCTACAGTGTCCCTAAACCAATGACTGCTTTAGTAGG-----AGA 565
 DB 243 AsnPheSerTyrSerProLeuLeuArgGluPheThrLysAlaThrAsnValArgLeuArg 262
 QY 566 TTTTTCGCCAGCATCAACAAATTA---AATCTCATTACACAACTTGGAGGCCCTTA 622
 DB 263 PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp----- 280
 QY 623 CCATCGGTAGTCAACGGGTGCTTTCCCTCGGATCAACTCTTCCAGGACCAACAAA 682
 DB 281 ProThrValThrArgArgTyrTyr-----SerIleLysAsp 293
 QY 683 ATAACCTTTGGTGGTACTGCTGACTGCTTCCAGTGGGACTTTTGC----- 730
 DB 294 IleSerIleGlyGlyArgCysValCysHisGlyHisAlaAspAlaCysAspAlaLysAsp 313
 QY 731 -----AACAACTGC-----AATGT 745
 DB 314 ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyGlyThrCys 333
 QY 746 AATAATTGTCGAACAACTTGCATCATGATTAATGACGGTTTAAAGCCATTAAAGCA--- 802
 DB 334 AspArgCysCysProGlyPheAsnGln-----GlnProTrpLysProAlaThrAlaAsn 351
 QY 803 -----TGCTCTGCTAGA-----AAT 817
 DB 352 SerAlaAsnGluCysGlnSerCysAsnCysTyrGlyHisAlaThrAspCysTyrTrpAsp 371
 QY 818 CCAGAA-----GCTTTCCAGCCAAAATTTGGAAGGGCAATTTGGCAAT 862
 DB 372 ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyrGlnGlyGly 391
 QY 863 GTC-----AAGCCCGCAGCAACAAAGGGTGCACTGCAGGAGGTGCGCTG 913
 DB 392 ValCysIleAspCysGlnHisThrThrGlyValAsnCysGluArg-----CysLeu 409
 QY 914 AAGATTACTGGAGTGTATGAGCCCAATATGTTCTTCTTATTTGCAAAATGCAT 973
 DB 410 ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgArgCys 429
 QY 974 GGTTCACAAAT-----TATGAAGAAAGCCAGAACGACACTA 1015
 DB 430 AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys 447
 QY 1016 ATGAGCATGCCAACTAC-----ATGCAGACTGGAGGTTGGAGGC 1057
 DB 448 TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly 467
 QY 1058 AGCCATTACCTGCCACCAACGAAATTTTCAGGACTTCCAGATTCAGTCACATAGGCGG 1117
 DB 468 PheProSerCysTyrProThr-----ProSerSerAsnAspThrArg 482
 QY 1118 CCTTCCTCATGCTCTCTGGAGGTGGTGGAGGCCACATGCGCTGCTGCTGCTCAG 1177
 DB 483 GluGlnValLeuProAlaGlyGlnIleValAsnCysAspCysSerAlaAlaGlyThrGln 502
 QY 1178 GGAGAAGAGCGGGAAGAACACTGCTCCAAAGTGCCTGGCAGACAGATGATCTGGAG 1237
 DB 503 GlyAsnAlaCysArgLysAsp-----Pro 510
 QY 1238 GAATTTGGAAGTGCTTATCA-----CAGATTCTCCACTCAGTTTAAATCT 1285
 DB 511 ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla 530
 QY 1286 AAGGAGTTGAAATGGAGTAGACTATAAGTGTGAATGC 1324
 DB 531 ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys 543

RESULT 4

BAR3_CHITE
 ID BAR3_CHITE STANDARD; PRT; 1700 AA.
 AC Q03376;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Balbiani ring protein 3 precursor.
 GN BR3.
 OS Chironomus tentans (Widge).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
 OC Chironomidae; Chironominae; Chironomus.
 OX NCBI_TaxID=7153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RX MEDLINE=90172404; PubMed=1689777;
 RA Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
 RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
 repetitive structure split by many introns.";
 RL J. Mol. Biol. 211:331-349(1990).
 CC !- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
 CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
 CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
 CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
 CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
 CC !- SUBCELLULAR LOCATION: Secreted.
 CC !- TISSUE SPECIFICITY: SALIVARY GLAND.
 CC !- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X52263; CAA36506.1; -;
 DR PIR; S08167; S08167.
 DR HSP; P15358; ISKZ.
 DR InterPro; IPR004153; CXXC-repeat.
 DR Pfam; PF03128; CXXC; 71.
 KW Repeat; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.
 SQ SEQUENCE 1700 AA; 186145 MW; 3420282852B0815 CRC64;
 Alignment Scores:
 Pred. No.: 0.0634 Length: 1700
 Score: 123.00 Matches: 87
 Percent Similarity: 31.73% Conservative: 45
 Best Local Similarity: 20.91% Mismatches: 146
 Query Match: 3.22% Indels: 139
 DB: 1 Gaps: 23
 US-09-743-237-3 (1-2134) x BAR3_CHITE (1-1700)
 QY 701 TGTGACTGCTTTGCCAGTGGGAC-----TTTTGC---AAC 733
 DB 198 CysSerCysGluCysLysGlyAspGlyLysCysGlnGlySerLysIleTrpCysLysAsn 217
 QY 734 AACTGCAATTGT-----AATAATTGTTGCCAACAACTTGCAT 769
 DB 218 AsnCysArgCysIleCysProThrAlaGluProAlaGlyCysSerAlaProLeuLys 237
 QY 770 CATGATATTGAAGCGTTTAAAGCCATTAAAGCATGTTGGTAGAAATCCAGAAGCTTTC 829
 DB 238 TrpAspAspAsp-----LysCysSerCysAlaCysProAlaLysMetGluGluLysLys 255
 QY 830 CAGCCAAAATTTGGGAAGGCCCAATTTGGCATGTCAACCCAGCAGCAACAAAGGTCG 889
 DB 256 GluLysCysValGluSerGlyLysIleTrpAsn-----ProAsnThrCysGluCys 272

```

Oy 890 AACTGCAGGAGGTGAGCTGCCTGAAGAAT-----TACTGC 925
    |||  |||  |||
Db 273 GlyCysAlaGlnLeuAsnCysProAspAsnLysLysAlaAsnLysGluThrCysGlnCys 292
    |||  |||  |||
Oy 926 GAGTGCATGAG-----GCCCAATATGTTGTTCTTCTATTGCAAA 967
    |||  |||  |||
Db 293 GluCysLysGluValLysCysAsnGlyGlnValPheCysLysAspSerCysSer 312
    |||  |||  |||
Oy 968 TGCATTGGTTGCAAAATATGAGAA-----AGCCAGAGCAAGAAACACA 1012
    |||  |||  |||
Db 313 CysVal---CysProGlyGlyAspLysAspLysThrCysThrAlaProGlnValTyAsp 331
    |||  |||  |||
Oy 1013 CTAATGAGCTGCAAACTACATCAGACGTGAGGTTTGAAGAGCACCATTACTCTCCCA 1072
    |||  |||  |||
Db 332 GlyVal-AlaCysSerCysCysProValAsnMetGlnLysProAlaAspGlyCysPr 351
    |||  |||  |||
Oy 1073 CCAACGAAATTTTCAGACTTCCAGATTC-----AGTCACGATAG 1113
    |||  |||  |||
Db 351 oArgProGlnLysTrpAspLysGluGluCysArgCysGluCysProValLysHisAspCy 371
    |||  |||  |||
Oy 1114 GCGGCTTCCTCATGCTCCTCGGAGGTGGTGAGGCACATGCCCTGCCTGCTGC 1173
    |||  |||  |||
Db 371 sLysAsnGlyLysVal-----TrpAsp-----GluThrIleCysGlnCysIleCysPr 387
    |||  |||  |||
Oy 1174 TCAG-----GGAGAGAGGCGCGAG----- 1192
    |||  |||  |||
Db 387 oArgAspAlaProValCysThrAlaGlyLysGluArgCysGlyGluSerCysGluCysLy 407
    |||  |||  |||
Oy 1193 -----AAGAACACTGCTCCAAG----- 1210
    |||  |||  |||
Db 407 sCysIleAsnArgGluProLysGluGlyCysAlaLysProLeuValTrpAsnGluAsnTh 427
    |||  |||  |||
Oy 1211 -----TGCTGCAGAGCAGATGATCTG-----GA 1236
    |||  |||  |||
Db 427 rCysLysCysValCysProAlaAspLysGlnMetSerProGlyGlyCysGlySerGlyLy 447
    |||  |||  |||
Oy 1237 GGAATTTGGAAGTGCTTATCACAGATTCCTCCACTGAGTTTAAATCTAAG---GGATT 1293
    |||  |||  |||
Db 447 sSerPheAsnLysLeuThrCysGlnCysGluCysAspGlnSerAlaSerLysCysGlyLe 467
    |||  |||  |||
Oy 1294 GAAATG-----GAGTAGAGTATAAGTGTGAATGCATGTTGTTTGTCTAGCTAG 1347
    |||  |||  |||
Db 467 uLysArgTrpAsnAlaAspThrCysLysCysGluCys----- 479
    |||  |||  |||
Oy 1348 AAATCTAGTTTAGAAGGATGTTAGGGGAACATGAGGTGCTGTCAGCAACAAC 1407
    |||  |||  |||
Db 480 -----GlnPr 481
    |||  |||  |||
Oy 1408 AGGCTCCCTGCATCCTCGGCGCCAGGAGTTTACTCA----- 1445
    |||  |||  |||
Db 481 oGlyMetProProGlu---GlyCysGlyLys-GlnThrTrpIleSerAspLysCysLysC 500
    |||  |||  |||
Oy 1446 --GAGCTCTCGAAGATGTGCAACCCATGCCCTTTCTGAGGAGTGCATGGCCCTGA 1503
    |||  |||  |||
Db 500 ysGluCysSerProThrIleThrCysGlnAlaProGlnIleLeuAspLeuAsnThrCysG 520
    |||  |||  |||
Oy 1504 GCATTGTTGT-----CTGGCCAGAGAGAGAGCTGGGTCCCATGAGTCCT 1551
    |||  |||  |||
Db 520 luCysLysCysProValAsnMetLeuAlaGlnLysGluLysCysLysSerProArgGlnT 540
    |||  |||  |||
Oy 1552 GG-----GAGAGTGTCTGAGGCGGCGGAGGAGCAGAGCCCTCGGAGAGCT 1602
    |||  |||  |||
Db 540 rPThrAspSerLysCysLeuCysGluCysSerThrThrProAlaThrCysGluGlyLysG 560
    |||  |||  |||
Oy 1603 CACTCTGGTGCATCTTCCCTCAGAGAATGTTCTCTGGAGGC 1646
    |||  |||  |||
Db 560 lnThrTrpCysGlyGluAlaCysGlnCysIleCysProGlyGly 574
    |||  |||  |||

```

RESULT 5
 AHM2_ARATH
 ID AHM2_ARATH
 AC O64474;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Potential cadmium/zinc-transporting ATPase 2 (EC 3.6.3.3)
 DE (EC 3.6.3.5)
 GN HMA4 OR ATZG19110 OR T20K24.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Root;
 RA Krijger G.C., Baccarini P.J., Williams L.E.;
 RT "Identification of ATHMA4, a putative heavy metal P-type ATPase in
 Arabidopsis";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Rinning C.M., Koo H.L.,
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana";
 RL Nature 402:761-768 (1999).
 CC -!- FUNCTION: INVOLVED IN CADMIUM/ZINC TRANSPORT (POTENTIAL).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + Cd(2+)(In) -> ADP + phosphate +
 Cd(2+)(Out).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + Zn(2+)(In) -> ADP + phosphate +
 Zn(2+)(Out).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 ATPases). Subfamily 1B.
 CC -!- SIMILARITY: Contains 1 HMA domain.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ297264; CAC19544.1; -
 CC EMBL; AC002392; AAD12041.1; -
 DR PIR; F84572; F84572.
 DR InterPro; IPR006416; ATPase-1B_hvy.
 DR InterPro; IPR001757; ATPase-1B_hvy.
 DR InterPro; IPR001756; Cu_ATPase.
 DR InterPro; IPR006404; Heavy_metal_ATPase.
 DR InterPro; IPR006121; Heavy_metal_transp.
 DR InterPro; IPR005834; Hydrolase.
 DR InterPro; IPR006191; Metal_bind.
 DR Pfam; PF00122; E1-E2_ATPase; 1.
 DR Pfam; PF00403; HMA; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR PRINTS; PR00943; CUATPASE.
 DR TIGRFAMs; TIGR01512; ATPase-1B2_Cd; 1.
 DR TIGRFAMs; TIGR01525; ATPase-1B_hvy; 1.
 DR TIGRFAMs; TIGR01494; ATPase-P-type; 2.
 DR PROSITE; PS00154; ATPase-1B2; 1.
 DR PROSITE; PS01047; HMA_1; FALSE_NEG.
 DR PROSITE; PS00846; HMA_2; 1.
 DR Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Magnesium;
 KW

Db	767	tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysCysGlyAspLysLys	787
Qy	746	-----AATAATTGTTGC----- :::	757
Db	787	sGluGluLysValLysProLeuValLysAspGlyCysCysSerGluLysThrArgLysSe	807
Qy	758	-----AACACTTGTCATCATGATAT :::	777
Db	807	rGluGlyAspMetValSerLeuSerSerCysLysLysSerHisValLysHisAspLe	827
Qy	778	TGAACGGTTTAAAGCCATTAAAGGCATCTCTTGGTAGAAATCCAGAACCTTCCAGCCAAA	837
Db	827	u-----LysMetLysGlySerGlyCysCysAlaSerLysAsnGluLysGlyLysGluVa	846
Qy	838	AATTGGGAAGCGCCAAATTGGCAATGCAAGCCCCAGCACACAAGGGTGCAACTCGAC	897
Db	846	lValAlaLysSerCysCysGluLysProLysGlnValGluSerValGlyAspCysLys	866
Qy	898	GAGTCAAGCTGCCTGAAG----- 	916
Db	866	sSerGlyHisCysGluLysLysLysGlnAlaGluAspIleValProValGlnIleI	886
Qy	917	---AATTACTGCGAGTGCTATGAGGCCCAATATATGTTCTCTTATTGCAATGCAT	972
Db	886	eGlyHisAlaLeuThrHisValGluIleGluLeuGlnThrLysGluThrCysLysThrSe	906
Qy	973	TGGTTGCAGAAAATTATCAAGAAAGCCAGACGAAGACACTAATGAGCATGCCAAAC--	1030
Db	906	rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuLeuSerGluAsnTh	926
Qy	1031	---TACATGCAGACTGGAGTTTG----- ::::	1062
Db	926	rProTyrLeuGluLysGlyValLeuIleLysAspGluGlyAsnCysLysSerGlySerGl	946
Qy	1063	TTACCTCCCAACGAATTTTCAGACATTCCAAGATTCCAGTACCATAGGCGGCCTTC	1122
Db	946	uAsnMetGlyThrValLysGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh	966
Qy	1123	CTCATGCATCTCTCGGAGGTGGTGGAGGCCACATGCGCCTGCTGCTCAGGGAGA	1182
Db	966	rGlyGluIleThr----- :::	975
Qy	1183	AGAGGCGGAGAAAGACACTGCTCCAAG----- ::::	1236
Db	975	uGluThrAspAspGlnAspCysSerSerGlyCysCysValAsnGluGlyThrValLysGl	995
Qy	1237	GGAATTTGGAAGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGATTTGAA	1296
Db	995	nSerPheAspGluLysLysHisSerValLeu----- :::	1012
Qy	1297	AATGGACTAGATATAAGTGTGAATGCATGTTGATTGTTGCTTAGTCTAGAAATCTCTA	1356
Db	1012	pMetGluThrGlyPheCysCysAspAlaLysLeuValCys----- :::	1025
Qy	1357	GTTTAGAAAGATGTTTAGGGGAACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCC	1416
Db	1025	----- :::	1025
Qy	1417	TGCATCCCTGGCCCCGAGGTACTTACTCAGAGCTCTCTGAAGATGTGGCAACCATGCC	1476
Db	1026	----- :::	1029
Qy	1477	CCTTTTCTGAGGAGTGCATGGCCTGAGCATTTGT--- 	1533
Db	1030	-----GluGlyGluValLysGluGlnCysArgLeuGluIleLysLysGluGlu--- 	1045
Qy	1534	TTGGGTTCCCATGACTCTCTGAGGAGTGTCTGACGGGCGGAGGAGCAGACAGGCCCTG	1593
Db	1046	-----HisCysLysSerGlyCysCysGlyGluGluIleGlnTh :::	1058
Qy	1594	CGGAGAGCTCACTCTGGTGCAGTCTTCTCTCAGAGAAATGT----- :::	1647
Db	1058	rGlyGluIleThrLeuValSerGluGluGluThrGluSerThrAsnCysSerThrGlyC	1078

QY 1648 GC 1649
 II
 Db 1078 ys 1078

RESULT 6
 POLG_HCVJA STANDARD; PRT; 3010 AA.

AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9108550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RA "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D90208; BAA14233.1; -;
 DR PIR; A39253; GNWVCU.
 DR HSP; P26663; IJXP.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.

InterPro; IPR001490; HCV_NS4b.
 InterPro; IPR002868; HCV_NS5a.
 InterPro; IPR002166; HCV_RdRP.
 InterPro; IPR001650; Helicase_C.
 InterPro; IPR007095; RNA_pol_DS_Ps.
 InterPro; IPR007094; RNA_pol_PSVir.
 Pfam; PF01543; HCV_capsid; 1.
 Pfam; PF01542; HCV_core; 1.
 Pfam; PF01539; HCV_env; 1.
 Pfam; PF01560; HCV_NS1; 1.
 Pfam; PF01538; HCV_NS2; 1.
 Pfam; PF02907; HCV_NS3; 1.
 Pfam; PF01906; HCV_NS4a; 1.
 Pfam; PF01001; HCV_NS4b; 1.
 Pfam; PF01506; HCV_NS5a; 1.
 Pfam; PF00271; helicase_C; 1.
 Pfam; PF00998; Viral_RdRP; 1.
 ProDom; PD186062; HCV_NS1; 1.
 SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3011 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Alignment Scores:
 Pred. No.: 0.141 Length: 3010
 Score: 119.50 Matches: 131
 Percent Similarity: 32.74% Conservative: 72
 Best Local Similarity: 21.13% Mismatches: 237
 Query Match: 3.12% Indels: 180
 DB: 1 Gaps: 35

US-09-743-237-3 (1-2134) x POLG_HCVJA (1-3010)
 QY 65 CCCCAGGATCCAGAGCTCAGCGCTGGAGAGCTCGCGCTCTCTGAGCGC----- 115
 Db 1217 ProAlaValProGlnThrPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLys 1236

QY 116 -----CCCGACGCGCGCGCTGCAACCTGCACTTCCTGCTCCTGCTAC 163
 Db 1237 SerThrLysValProAlaAlaTyrAlaAlaGln---GlyTyrLysValLeuValLeuAsn 1255
 QY 164 CCGCAGCCGCGCGCGGCTGTTTCCCTCGCGCGCGCTG-----GGTCTCGG 214
 Db 1256 ---ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGly 1273
 QY 215 AAGGAGCTCCACCCGCGCGCTGCGCATGATCCAGTTCAGAAATCAAGTAAGCAGGTGGT 274
 Db 1274 IleGluProAsnIleArgThrGlyVal-----Arg 1283
 QY 275 ACTACTACAAGTAATAATCCGGAAGAACTTTCAGAAATCTTCTGTCAGAAATCC 334
 Db 1284 ThrIleThrThrGlyPro-----IleThrTyrSerThr 1295
 QY 335 TGTTCAAGTTCCCGCGCGCTGCGAGGATGCTCCTGCTGTTCTCTTAAGAA 394
 Db 1296 TyrCysLysPhe-----LeuAlaAspGlyGlyCysSerGlyGlyAlaTyr 1310
 QY 395 GATTCCAAACCAATGGTGATATGCCAATTTGAAGGGGCGCACAAATGCTATGTATA--- 451
 Db 1311 Asp-----IleIleIleCysAspGluCysHisSerThrAspSerThrThrIleLeu 1327
 QY 452 -----GACAAATCTAGACACAGAGAACTAAAGACACTCCATTTGGTT 493
 Db 1328 GlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValLeuAla 1347
 QY 494 -----CCTCAGTATCAAGATCAAAATAT 517
 Db 1348 ThrAlaThrProGlySerIleThrValProHisProAsnIleGluGluValAlaLeu 1367
 QY 518 TATCTACAGTCAGATGCTCCT-----AAACCAATGACTCTTTAGTAGGG 562
 Db 1368 SerAsnThrGlyGluIleProPheTyrGlyLysAlaIleProIleGluAlaIleLysGly 1387
 QY 563 AGA-----TTTTGCCAGCATCAACAAATTAATCTCATTACACAACT 610
 Db 1388 GlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAlaLysLeu 1407
 QY 611 GAGGA---GCCTTACCATCGTA-----GTCAACGGTCTGCTTCC 652
 Db 1408 ThrGlyLeuGlyLeuAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIlePro 1427
 QY 653 ---TCGGATCAACTCTCCAGGACCCACAAATAAATCTTGGCTGGGTACTGTGCTGCTG 709
 Db 1428 ThrSerGlyAspValValValAlaThrAlaThrAlaLeuMetThrGlyPhe----- 1444
 QY 710 TTGCGCAGTGGGACTTTTGCAACACTGCAATGTAATTAATTTGTCACAACTG--- 766
 Db 1445 ---ThrGlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrVal 1462
 QY 767 -----CATCATGATATTGACCGGTTTAAGCCATTAAAGCATGTCTT 808
 Db 1463 AspPheSerLeuAspProThrPheThrIleGluThrThrThrLeuProGlnAspAlaVal 1482
 QY 809 GGTAGAAATCCAGAGCTTTCACCAAAATTTGGAGGCGCAATTTGGC----- 859
 Db 1483 SerArgAlaGlnArg-----ArgGlyArgThrGlyArgGlyArgSerGlyIleTyrArg 1500
 QY 860 AATGTCAAGCCCGACACACAAAGGTTGCAACTGCAGGAGGTGCGCTGCGTGAAGAT 919
 Db 1501 PheValThrProGlyGlu-----ArgProSerGlyMetPheAspSer 1514
 QY 920 -----TACTGGCAGTGTATGAGCCCAATTTATGTGTTCTTCTATTTCGAAATGCATT 973
 Db 1515 SerValLeuCysGluCysTyrAspAla----- 1523
 QY 974 GGTTCGCAAAATTTATGAGAAAGCCAGACAAAGACACTAATGAGCATGCCAACTAC 1033
 Db 1524 GlyCysAlaTyrTyrGluLeuThrPro---AlaGluThrSerValArgLeuArgAlaTyr 1542
 QY 1034 ATGCAGACTGGAGTTTGAA----- 1054

Db 1543 LeuAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTyrGluSerValPhe 1562
 QY 1055 -----GGCAGCATTACCTGCGCCACCAACCAAAATTTTCAGGA----- 1090
 Db 1563 ThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsn 1582
 QY 1091 CTTCCAAGATTTCAGTC-----ACGATAGCGCGCTTCCTCATGTCATCTCCTCGG 1138
 Db 1583 LeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProPro 1602
 QY 1139 GAGGTGG-----TGGAGGCCACATCGCGCTGCTGCTGCTCAGGAGAAAGAGCC 1189
 Db 1602 oSerTyrAspGlnMetTyrLys-----CysLeuIleArgLeuLysProth 1617
 QY 1190 GAGAAAGAACACTGCTGCTCAAGTGCCTGCGAGACAGATGATCCTGGAGGAATTTGAAGG 1249
 Db 1617 rLeuHisGlyProThrProLeuLeuTyrArgLeu-----G 1629
 QY 1250 TGCTTATCACAGATTTCACACTGAGT----- 1277
 Db 1629 yAlaValGlnAsnGluValThrLeuThrHisProIleThrLysTyrIleMetAlaCysMe 1649
 QY 1278 -----TTAAATCTAAGGATTGAAATGGAGTAGAGTAGAGTATAAAGTGTGAATGCATG 1327
 Db 1649 tSerAlaAspLeuGluValValThrSerThrTyrValLeuValGlyValLeuAlaAl 1669
 QY 1328 TTGATTTTCTTACTAGAAATCTCTAGTTTGAAGAGGATGTTTAGGGGACATGAGG 1387
 Db 1669 aLeuAlaTyrCysLeuThrThrGlySerValValIleVal-----G 1684
 QY 1388 CTGGCTCTGCAACAACACAGGCTCCCTGCTGCTGCGCCAGGAGTTTACTCAGA 1447
 Db 1684 yArgIleIleLeuSerGlyArgProAlaValIleProAspArgGluValLeuTyrGlnG 1704
 QY 1448 GCTCTCT-----GAAGATGTGGCAACCATGCCCTTTCTGAGGAGGTGCATGGCT 1501
 Db 1704 uPheAspGluMetGluGluCysAlaSerHisLeuProTyrIleGluGln-----GlyMe 1722
 QY 1502 GAGCATTTGTTCTGCTGCGCCAGGAGAGAGCTGGTTCCCATAGTCTCTGGGAGAGTGT 1561
 Db 1722 tGlnLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-----LeuGlnThrAl 1739
 QY 1562 CTGCAAGCGCGGAGGCGGAGAGGCGCTCGGAGAGCTCACTCTGTGCTGACTCTT 1619
 Db 1739 aThrLysGlnAlaGluAla---AlaAlaProValValGluSerLysTyrArgAlaLeu 1757

RESULT 7

YLK3_CABEL
 ID YLK3_CABEL STANDARD; PRT; 1576 AA.
 AC P41951; Q950P7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative serine/threonine-protein kinase D1044.3 in chromosome III
 DE (EC 2.7.1.1-).
 OS Caenorhabditis elegans.
 GN Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Pauley A., Waterston R.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RA Waterston R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=a;

CC Name-b: IsoId=P41951-1; Sequence=Displayed;

CC IsoId=P41951-2; Sequence=VSP_004900, VSP_004901;

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC STRONG, TO ZC84.1.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: U00065; AAK68286.1; -

DR EMBL: U00065; AAL27237.1; -

DR HSSP: Q63450; 1A06.

DR WormPep; D1044.3a; CE27894.

DR WormPep; D1044.3b; CE29743.

DR InterPro; IPR006149; EB_region.

DR InterPro; IPR000719; Ser_kinase.

DR InterPro; IPR002290; Ser_thr_kinase.

DR InterPro; IPR006150; Worm_repeat_1.

DR Pfam; PF01683; EB; 8.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; prot_kinase; 1.

DR SMART; SM00220; S_TKC; 1.

DR SMART; SM00289; WRL; 8.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR Hypothetical protein; Transferase; Serine/threonine-protein kinase;

DR KW ATP-binding; Alternative splicing.

DR FT DOMAIN 431 703 PROTEIN KINASE.

DR FT NP_BIND 437 445 ATP (BY SIMILARITY).

DR FT BINDING 461 461 ATP (BY SIMILARITY).

DR FT ACT_SITE 569 569 Missing (in isoform b).

DR FT VARSP_LIC 1 904 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569 GLUTPFSNSNPQUTCPQDACSCYSSGSRFGTICQYA

DR FT VARSP_LIC 1 904 STYNNIYICSTNTQ (in isoform b).

DR FT VARSP_LIC 905 1012 /FTID=VSP_004901.

DR FT VARSP_LIC 905 1012 /FTID=VSP_004900.

DR FT VARSP_LIC 905 1012 SPECTOMLGFDFRPLVQSHINGYQKDKVTKITYEOLS

DR FT ACT_SITE 569 569 ACTCLIAENPAKHPHTRAVVILTRDLVLGVQYNNLV

DR FT NP_BIND 437 445 PTITTYVVIROIHVSLAAILVYVE -> MEETCSPEKE

DR FT BINDING 461 461 SNISFIMHKLKRVPPIMICLFLLLQIIEVISVSCPP

DR FT ACT_SITE 569 569

QY 967 ATGCATT-----GTTGCAAAATATGAAGAAACCCAGAACG 1005
 :|||||
 Db 1133 nCysIleThrLeuAlaSerProGlyGlnAsnCysGlnThrSerSerGlnCysIleAspAs 1153
 :|||:|
 QY 1006 AAAGACACTAATGAGCATG-----CCAAACTACATGCAGACTGGAGGTTT 1050
 :|||:|
 Db 1153 nSerValCysMetAsnGlnMetCysThrCysAsnAsnAsnTyrArgLeuValTyrGly-- 1172
 :|||:|
 QY 1051 GGAAGGACGACATTACCTGCCACCAACGAAATTTTCAGGACTTCCAAGATTCAAGTCACGA 1110
 :|||:|
 Db 1173 -----TyrCysValProIleThr-----SerSerIleCysGlnGlnThrGlnTh 1187
 :|||:|
 QY 1111 TAGGGCGCTTCCTCATCATCTCTCTGGAGGTGGAGCCACATGC----- 1159
 :|||:|
 Db 1187 rLeuValAsnAsnGlnCysValLeuLeuSerIleValGlyGluThrCysIleAlaAsnG1 1207
 :|||:|
 QY 1160 -----GCCCTGCTGCTGCTCAGGGAGAAGA 1185
 :|||:|
 Db 1207 nGlnCysValGlyGlyAlaMetCysAsnSerGlyThrCysGlnCysThrAsnGlyAlaTh 1227
 :|||:|
 QY 1186 GCGCGAGAAAGAACACTGC-----TCCAAGTCCCTGGCAGACGACATGAT 1230
 :|||:|
 Db 1227 rAlaMetTyrGlyThrCysIleSerSerSerSerSerSerSerSerSerSerSerSerSer 1247
 :|||:|
 QY 1231 CCTGGAGGAATTGGAAGTCTTATCATCAGATT 1264
 :|||:|
 Db 1247 rIleAsn-----GlyMetCysTyrAsnThrVal 1256
 :|||:|
 RESULT 8
 LMB2_RAT STANDARD; PRT: 1801 AA.
 ID LMB2_RAT
 AC P15800;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Laminin beta-2 chain precursor (S-laminin) (Laminin chain B3).
 GN LAMB2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Hunter D.D., Shah V., Merlie J.P., Sanes J.R.;
 RT "A laminin-like adhesive protein concentrated in the synaptic cleft
 of the neuromuscular junction.";
 RL Nature 338:229-234(1989).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 different polypeptide chains (alpha, beta, gamma), which are bound
 to each other by disulfide bonds into a cross-shaped molecule
 comprising one long and three short arms with globules at each
 end.
 CC THE BETA-2 CHAIN IS A SUBUNIT OF LAMININ-3 (S-LAMININ), LAMININ-4
 (S-MEROSIN), AND LAMININ-7 (KS-LAMININ).
 CC -!- CELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 COMPONENT). S-LAMININ IS CONCENTRATED IN THE SYNAPTIC
 CLEFT OF THE NEUROMUSCULAR JUNCTION.
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -!- SIMILARITY: Contains 13 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 1 laminin IV domain.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch.

 CC EMBL; X16563; CAA34561.1; -
 DR PIR; S03539; MMRTS.
 DR HSP; P02468; 1KLO.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001886; LamNT.
 DR Pfam; PF00053; laminin_EGF; 13.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 13.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 10.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
 DR GlycoProtet; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 35
 FT CHAIN 36 1801 LAMININ BETA-2 CHAIN.
 FT DOMAIN 36 283 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 286 349 LAMININ EGF-LIKE 1.
 FT DOMAIN 350 412 LAMININ EGF-LIKE 2.
 FT DOMAIN 413 472 LAMININ EGF-LIKE 3.
 FT DOMAIN 473 524 LAMININ EGF-LIKE 4.
 FT DOMAIN 525 555 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 556 785 LAMININ DOMAIN IV.
 FT DOMAIN 786 833 LAMININ EGF-LIKE 6.
 FT DOMAIN 834 879 LAMININ EGF-LIKE 7.
 FT DOMAIN 880 929 LAMININ EGF-LIKE 8.
 FT DOMAIN 930 988 LAMININ EGF-LIKE 9.
 FT DOMAIN 989 1040 LAMININ EGF-LIKE 10.
 FT DOMAIN 1041 1097 LAMININ EGF-LIKE 11.
 FT DOMAIN 1098 1145 LAMININ EGF-LIKE 12.
 FT DOMAIN 1146 1192 LAMININ EGF-LIKE 13.
 FT DOMAIN 1193 1412 DOMAIN II.
 FT DOMAIN 1413 1445 DOMAIN ALPHA.
 FT DOMAIN 1446 1801 DOMAIN I.
 FT DOMAIN 1259 1306 COILED COIL (POTENTIAL).
 FT DOMAIN 1475 1529 COILED COIL (POTENTIAL).
 FT DOMAIN 1576 1793 COILED COIL (POTENTIAL).
 FT DISULFID 286 295 BY SIMILARITY.
 FT DISULFID 288 313 BY SIMILARITY.
 FT DISULFID 315 324 BY SIMILARITY.
 FT DISULFID 327 347 BY SIMILARITY.
 FT DISULFID 350 359 BY SIMILARITY.
 FT DISULFID 352 377 BY SIMILARITY.
 FT DISULFID 380 389 BY SIMILARITY.
 FT DISULFID 392 410 BY SIMILARITY.
 FT DISULFID 413 426 BY SIMILARITY.
 FT DISULFID 415 441 BY SIMILARITY.
 FT DISULFID 443 452 BY SIMILARITY.
 FT DISULFID 455 470 BY SIMILARITY.
 FT DISULFID 473 487 BY SIMILARITY.
 FT DISULFID 475 494 BY SIMILARITY.
 FT DISULFID 496 505 BY SIMILARITY.
 FT DISULFID 508 522 BY SIMILARITY.
 FT DISULFID 786 798 BY SIMILARITY.
 FT DISULFID 788 805 BY SIMILARITY.
 FT DISULFID 807 816 BY SIMILARITY.
 FT DISULFID 819 831 BY SIMILARITY.
 FT DISULFID 834 846 BY SIMILARITY.
 FT DISULFID 836 853 BY SIMILARITY.
 FT DISULFID 855 864 BY SIMILARITY.
 FT DISULFID 867 877 BY SIMILARITY.
 FT DISULFID 880 889 BY SIMILARITY.
 FT DISULFID 882 896 BY SIMILARITY.
 FT DISULFID 899 908 BY SIMILARITY.
 FT DISULFID 911 927 BY SIMILARITY.

FT	DISULFID	930	946	BY SIMILARITY.
FT	DISULFID	932	957	BY SIMILARITY.
FT	DISULFID	959	968	BY SIMILARITY.
FT	DISULFID	971	986	BY SIMILARITY.
FT	DISULFID	989	1003	BY SIMILARITY.
FT	DISULFID	991	1010	BY SIMILARITY.
FT	DISULFID	1013	1022	BY SIMILARITY.
FT	DISULFID	1025	1038	BY SIMILARITY.
FT	DISULFID	1098	1110	BY SIMILARITY.
FT	DISULFID	1100	1117	BY SIMILARITY.
FT	DISULFID	1119	1128	BY SIMILARITY.
FT	DISULFID	1131	1143	BY SIMILARITY.
FT	DISULFID	1146	1158	BY SIMILARITY.
FT	DISULFID	1148	1165	BY SIMILARITY.
FT	DISULFID	1167	1176	BY SIMILARITY.
FT	DISULFID	1179	1190	BY SIMILARITY.
FT	DISULFID	1193	1193	INTERCHAIN (PROBABLE).
FT	DISULFID	1196	1196	INTERCHAIN (PROBABLE).
FT	DISULFID	1800	1800	INTERCHAIN (PROBABLE).
FT	CARBOHYD	251	251	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	371	371	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1088	1088	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1252	1252	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1311	1311	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1351	1351	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1502	1502	N-LINKED (GLCNAC. . .)
SQ	SEQUENCE	1801 AA; 196473 MW; 97AEF32F8F31FA75 CRC64;		

Alignment Scores:				
Pred. No.:	0.149	Length:	1801	
Score:	118.50	Matches:	110	
Percent Similarity:	27.02%	Conservative:	54	
Best Local Similarity:	18.12%	Mismatches:	193	
Query Match:	3.10%	Indels:	250	
DB:	1	Gaps:	29	

US-09-743-237-3 (1-2134) x LMB2_RAT (1-1801)				
Qy	76	CAGAGCTCAGCGCTGGAGGACGTCGCGTCTCGCAGCGCCGCGCGCCGCTGCA	135	
Db	633	GlnValProGluGlnTTPAlaGluLeuValGlnArgProGlyProValSer	652	
Qy	136	ACGTGCACTTCTGCTGCTGCTGCTACCGGGACCGCGCGCGGCTTTTGGCCCT	195	
Db	653	AlaHisSerProCysGlyHisValLeuProArg	663	
Qy	196	GGGCGCTGGTCTCGGAGGAGCTCCACCGCGCGCTCCGATGATC	246	
Db	664	-----AspAspArgIleGlnGlyMetLeuHisProAsnThrArgValLeuValPhePro	681	
Qy	247	---CCAGTTGAATCAAGGTAAGCAGGTGGTACTACTACAAG-----TAATAA	291	
Db	682	ArgProValCysLeuGluProGlyLeuSerTyrlsLeuLysLeuLysLeuThrGlyThr	701	
Qy	292	TCCGGAAGAACAAC-----TTTGCAGAAATCTTTGGCTCAGAAATCCTGTTG	345	
Db	702	GlyGlyArgAlaHisProGluThrProTyrlsSerGlyLeuLeuLeuLeuAspSerLeu	721	
Qy	346	CCATGTCTCCAGAACTAGAGGATGCTGCTGCTTCTTAAGAAAGATCCACCC	405	
Db	722	ValLeuGlnProHisVal-----LeuMetLeu-----GluMetPheSerGly	735	
Qy	406	AATGGTGATATGCCAATTTAAAGGGGCACACAAATGCTATGATAGACAAATCTAGAAC	465	
Db	736	GlyAspAlaAlaLeuGluArgArgThrThrPheGluArgTyrlsArg-----	751	
Qy	466	AAGAACTAAAGCACTCCATTTTGGTTCCTAGTATCAAGATCAAAATAATATCTACA	525	
Db	751	-----	751	
Qy	526	GTCAGATGTCCTAAACCAATGACTGCTTTAGTAGGAGGATTTTGGCAGCATCAACAA	585	
Db	752	-----CysHisGluGlyLeu-MetProSerLysThrPr	763	

Qy	586	ATTAATCTCATTACACAACAACCTTGAGGAGCCTTACCATCGGTAGTCAACGGTCTGC	645	
Db	763	oLeuSer-----GluAlaCysValProLeuLeuSerAlaSerSe	777	
Qy	646	TTTCCCTCCGGATCAACTCTTCAGGAGCCACCACAAAATAACTTTGGCTGGTACTGTGA	705	
Db	777	rLeuValTyrlsAsnGlyAlaLeuPro-----CysG1	787	
Qy	706	CTGCTTTGCCAGTGGGACCTTTTCCACAACACTGCAATTGTAATAATTGTTCCACAACCTT	765	
Db	787	nCysAspProGlnGlySerLeuSerSerGluCysAsnProHisGly	802	
Qy	766	GCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAATCCAGAAGC	825	
Db	803	-----GlyGlnCysArgCysLysProGlyValValGlyArgCysAspAl	818	
Qy	826	TTTCCAGCCAAAATTTGGGAGGGCCAAATTTGGGCAATGTCAAGCCCGCAGCACACAAGG	885	
Db	818	a-----CysAlaThrGlyTyrlsGly---PheGlyProAlaGlyCysGlnAl	833	
Qy	886	GTCAACTGCAGGAGGTTCAGGCTCCCTGAAGAATTACTGCGAGTGCTATGAGGCCCAAT	945	
Db	833	aCysGlnCysSerProAspGlyAlaLeuSerAlaLeuCysGlnGlyThrSerGlyGlnC	853	
Qy	946	TATGTGTCTTCTCT	958	
Db	853	sLeuCysArgThrGlyAlaPheGlyLeuArgCysAspHisCysGlnArgGlnTrpG1	873	
Qy	959	-----ATTTCGAAA-----	967	
Db	873	yPheProAsnCysArgProCysValCysAsnGlyArgAlaAspGluCysAspAlaHisTh	893	
Qy	968	-----TGCATTGTTGCAAAATTTATGAAGAAAGCCAGAACGAAAGACACTAATGAG	1020	
Db	893	rGlyAlaCysLeuGlyCysArgAspTyrls	902	
Qy	1021	CATGCCAAACTACATGCAGACTGGAGGT-----TTGGAAGCGACCCA	1062	
Db	903	-----ThrGlyGlyGluHisCysGluArgCysIleAlaGlyPheHi	916	
Qy	1063	TTACCTCCCAACCAAGAAATTTTCAGGACTTCCAGAGTTTCCAGTACCATGAGCGGCCTTC	1122	
Db	916	sGlyAspProArgLeuProTyrlsGlyGlnCysArg	928	
Qy	1123	CTCATGATCTCTCGGAGGTGGTGGAGGCCACATGCGCTGCTGCTGCTCAGGAGA	1182	
Db	929	-----ProCysProCys-----ProGluGlyPr	936	
Qy	1183	AGAGCGGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGACGATGATCTTGGAGGAAT	1242	
Db	936	oGlySerGlnArgHisPheAlaThrSerCysHisArgAsp	949	
Qy	1243	TGGAAGTGTCTTATCAGATTTCCACATGACTTTAAATCTAAGGATTTGAAATGGA	1302	
Db	950	-GlyTyrlsSerGlnGlnIleValCysHisCysArgAlaGlyTyrlsThrGlyLeuArgCysG1	969	
Qy	1303	GTAGAGTATAAG	1317	
Db	969	uAlaCysAlaProGlyHisPheGlyAspProSerLysProGlyGlyArgCysGlnLeuCy	989	
Qy	1318	TGAATGATGTTGATTTTGTCTTAGTCTAGAAAATCTCTAGTTTAAAGGATTTTAGG	1377	
Db	989	sGluCysSer-----	993	
Qy	1378	GAACATGAGGCTGCTGTCAGCAACACAGGCTCCCTGTCATCCCTGGGCCCA	1432	
Db	993	yAsnIle-----AspProThrAspProGlyAlaCysAspProHisTh	1007	
Qy	1433	-GGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCGCCCTTTTCTCAGGAGG	1491	
Db	1007	rGlyGlnCysLeuArgCysLeuHisThrGluGlyPro-----	1020	


```
QY 1492 TGCATGGCGCTGAGCATTTGTTCTCTGCGCCAGAGGAGAGCTTGGTTCCCATAGTCCT 1551
Db 1021 -----HIScys-----
QY 1552 GGGAGAGTCTTCGACGGCGGGGGAGGAGCAGC-----AGCCCTCGGGAGAGCTCAC 1605
Db 1023 -GlyHIScysLysProGlyPheHISglyGlnAlaLaArgGlnSerCysHISArgCysH 1042
QY 1606 T-----CTGGTCGACTCTCTCTCTCAGAGAATGTTGCTCTGGAGGCTCTCTGCATGAA 1659
Db 1042 rcysasnLeuLeuGlyThrAspProGlnArg-----CysProSe 1055
QY 1660 AACCCCTAATGGTTTCTTGT 1678
Db 1055 rThrAspLeuCysHIScys 1061

RESULT 9
ZAN_MOUSE
ID ZAN_MOUSE STANDARD; PRT: 5376 AA.
AC O88799; O08647;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT "Species diversity in the structure of zonadhesin, a sperm-specific
RT membrane protein containing multiple cell adhesion molecule-like
RT domains.";
RL J. Biol. Chem. 273:3415-3421(1998).
RN [2]
RP SEQUENCE OF 4864-5376 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD.
CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -1- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -1- SIMILARITY: Contains 3 MAM domains.
CC -1- SIMILARITY: Contains 25 VWFD domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; U97068; AAC36680.1; -.
```

```
DR EMBL; U83190; AAC53125.1; -.
DR PIR; T42215; T42215.
DR MGD; MGI:106656; Zan.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003645; FOLN.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR003328; TILA_Cysrich.
DR InterPro; IPR001007; WVF_C.
DR InterPro; IPR001846; WVF_D.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILA; 25.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FOLN; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00060; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 5376 ZONADHESIN.
FT DOMAIN 5311 5337 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5338 5376 POTENTIAL.
FT DOMAIN 45 210 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 215 374 MAM 1.
FT DOMAIN 377 542 MAM 2.
FT DOMAIN 547 1170 MAM 3.
FT DOMAIN 1171 1280 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT DOMAIN 1281 1669 (MUCIN-LIKE DOMAIN).
FT DOMAIN 1670 2056 VWFD 1 (PARTIAL).
FT DOMAIN 2057 2459 VWFD 2.
FT DOMAIN 2460 2579 VWFD 3.
FT DOMAIN 2580 2699 VWFD 4.
FT DOMAIN 2700 2819 VWFD 5 (PARTIAL).
FT DOMAIN 2820 2939 VWFD 6 (PARTIAL).
FT DOMAIN 2940 3059 VWFD 7 (PARTIAL).
FT DOMAIN 3060 3179 VWFD 8 (PARTIAL).
FT DOMAIN 3180 3299 VWFD 9 (PARTIAL).
FT DOMAIN 3300 3416 VWFD 10 (PARTIAL).
FT DOMAIN 3417 3536 VWFD 11 (PARTIAL).
FT DOMAIN 3537 3656 VWFD 12 (PARTIAL).
FT DOMAIN 3657 3776 VWFD 13 (PARTIAL).
FT DOMAIN 3777 3892 VWFD 14 (PARTIAL).
FT DOMAIN 3893 4029 VWFD 15 (PARTIAL).
FT DOMAIN 4029 4148 VWFD 16 (PARTIAL).
FT DOMAIN 4149 4263 VWFD 17 (PARTIAL).
FT DOMAIN 4264 4383 VWFD 18 (PARTIAL).
FT DOMAIN 4384 4503 VWFD 19 (PARTIAL).
FT DOMAIN 4504 4623 VWFD 20 (PARTIAL).
FT DOMAIN 4624 4743 VWFD 21 (PARTIAL).
FT DOMAIN 4744 4863 VWFD 22 (PARTIAL).
FT DOMAIN 4864 5261 VWFD 23 (PARTIAL).
FT DOMAIN 5259 5295 VWFD 24 (PARTIAL).
FT DISULFID 5263 5274 VWFD 25.
FT DISULFID 5268 5283 EGF-LIKE.
FT DISULFID 5285 5294 BY SIMILARITY.
FT DISULFID 5295 5304 BY SIMILARITY.
FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1216 1216 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1314 1314 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1814 1814 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1933 1933 N-LINKED (GLCNAC. .) (POTENTIAL).
```


Db 3945 sValLeuGln---ThrHisCysAspCysLysAspAlaGluGlyLeuValProAlaG1 3964
QY 1291 ATTGAATAGGAGTAGAGTATAAGTGTGAATGCTGTTGATTTGCTCTAGTAGAAA 1350
Db 3964 yLys-ThrTrpThrSerLysAspCysThrGlnSerCys---AlaCysThrGlyGlyAlaV 3983
QY 1351 TCTCTAGTTTGAAGAAGATGTTTAGGGGAACATGAGCTGCTGCACCAACACAGG 1410
Db 3983 aIcGln-----CysGlnAsnPheGlnC 3990
QY 1411 CTCCTCGTCCCTGGGCCCCAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCC 1470
Db 3990 yProLeu-----GlyThrCysLysAspSerGlyAspGlySerSera 4005
QY 1471 ATGCCCCCTTTCTGAGGAGTGCATGCGCTGAGCATTTCTGCTGCGCCAGAGAGAG 1530
Db 4005 sCys-----ThrLysIleHis-----LysGlyA 4013
QY 1531 AGCTTGGGTTCCCATAGTCTCTGGGAGAGTGTCTGCAGGCGGGAGGCGAGCA---- 1586
Db 4013 laMetGly-----AspGlyValLeuMetAlaGlyGlyIleArgAlaLeuG 4028
QY 1587 --GGCCCTCGGAGAGCTCACTCTGCTGCTGCTCTCTCAGAGAAATGTTGCTCTGGAG 1644
Db 4028 lncysProAlaHisSerHisPheThrSerCysLeuPro-----SercysProProS 4045
QY 1645 GCTGCTCT 1652
Db 4045 erCysSer 4047
RESULT 10
POLG_HCV1 STANDARD; PRT; 3011 AA.
ID AC POLG_HCV1
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11104;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M62321; AAA45676.1; -
DR PIR: A39166; GNWVC3.
DR PDB: 1A1V; 16-FEB-99.
DR PDB: 1HEI; 25-NOV-98.
DR MEROPS; S29.001; -
DR MEROPS; U39.001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 POTENTIAL.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Alignment Scores:
 Pred. No.: 0.226 Length: 3011
 Score: 117.00 Matches: 91
 Percent Similarity: 35.49% Conservative: 46
 Best Local Similarity: 23.58% Mismatches: 138
 Query Match: 3.06% Indels: 111
 DB: 1 Gaps: 23

US-09-743-237-3 (1-2134) x POLG_HCV1 (1-3011)

QY 641 TCTGTTTCCCC---TCGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGG 697
 Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
 QY 698 TACTGTGACTGCTTGGCAGTGGGACCTTTTGCACAACTCAATTGCTAAATTTGCTGC 757
 Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
 QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
 Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
 QY 797 AAGGCATCTTGTGTAATAATCCAGAGCTTTCAGCCCAAAATAATGGGAAGGCCAATG 856
 Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
 QY 857 GGC-----AATGTCACGCCCCAGCACACAAAGGGTCAACTGCAGGAGTCAGGC 907
 Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
 QY 908 TGCCTGAAGAAT-----TACTGGCAGTGTCTATGAGGCCCAATAATATGTCTTCTATT 961
 Db 1511 MetPheAspSerSerValLeuCysGlyCysTyrAspAla----- 1523
 QY 962 TGCAATCCATGTTGTGCAAAATATGCAAGAACCCAGACCAAGAACACACTAATGAGC 1021
 Db 1524 -----GlyCysAlaTyrPyrGluLeuThrPro-----AlaGluThrThrValArg 1538
 QY 1022 ATGCCAAACTACATGCAGACTGGAGTTG----- 1051
 Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
 QY 1052 GAAGC-----AGCCATTACCTGCCACCAACCAAAATTT 1084
 Db 1559 GluGlyValPheThrGlyLeuThrHisTleAspAlaHisPheLeuSerGlnThrLysGln 1578
 QY 1085 TCAGGA-----CTTCAAGATTTCAGTC-----ACGATAGCGCGCTTCCTCA 1126
 Db 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 1598
 QY 1127 TGCATCTCCTGGAGGTGG-----TGAGGCCACATCGCTCGCTTGTCTCAG 1177
 Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1613
 QY 1178 GGAGAAGAGCGCGAGAAACACTGCTCCCAAGTGCCTGGCAGCAGACATGATCCTGGAG 1237
 Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
 QY 1238 GAATTGGAAGTGTTCATCAGATTCTCCACACTGAGT----- 1277
 Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1645
 QY 1278 -----TTAAATCTTAAGGATTGAAATGGAGTACAGATATAAG 1315
 Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 1665

QY 1316 TGTGAATGCA-----TGTTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAGCAT 1369
 Db 1665 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal----- 1683
 QY 1370 GTTTAGGGAACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCTGTCATCCCTGGGC 1429
 Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
 QY 1430 CCAGGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGCAACCCATGCCCTTTTC 1483
 Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1718
 QY 1484 TGAGGAGGTGTCAGCTGCTGAGCATTTTGTCTGCCGCCAGAGAGAGAGCTTGGGTTCCC 1543
 Db 1718 eGluGln-----GlyMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
 QY 1544 ATAGTCTGGGAGAGTGTCTGCAGGGCGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGTC 1603
 Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla----ProAlaValGlnTh 1752
 QY 1604 ACTCTGTCGACTCTT 1619
 Db 1752 rasntrpGlnLysLeu 1757

RESULT 11
 LMA5_MOUSE
 ID LMA5_MOUSE STANDARD; PRT; 3718 AA.
 AC Q61001; Q9JH06;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Laminin alpha-5 chain precursor.
 GN LMA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE OF 1-92 FROM N.A., AND SEQUENCE OF 41-46.
 RX MEDLINE=21818471; PubMed=11829758;
 RA Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
 RT "Complete sequence, recombinant analysis and binding to laminins and
 sulphated ligands of the N-terminal domains of laminin alpha3B and
 alpha5 chains.";
 RL Biochem. J. 362:213-221(2002).
 RN [2]
 RP SEQUENCE OF 84-3718 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=96081906; PubMed=7499364;
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
 expression in adult mouse tissues.";
 RL J. Biol. Chem. 270:28523-28526(1995).
 RN [3]
 RP REVISIONS.
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
 ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 different polypeptide chains (alpha, beta, gamma), which are bound
 to each other by disulfide bonds into a cross-shaped molecule
 comprising one long and three short arms with globules at each
 end.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 MEMBRANES (MAJOR COMPONENT).
 CC -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
 KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT

CC AND SKIN, EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -I- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC
CC -I- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC
CC -I- SIMILARITY: Contains 1 laminin N-terminal domain.
CC
CC -I- SIMILARITY: Contains 22 laminin EGF-like domains.
CC
CC -I- SIMILARITY: Contains 2 laminin IV domains.
CC
CC -I- SIMILARITY: Contains 5 laminin G-like domains.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcements>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL: AJ293593; CAB99255.1; -.
CC DR EMBL: U37501; AAC53430.1; -.
CC DR PIR: T10053; T10053.
CC DR HSSP: P02468; ITLE.
CC DR MGD; MGI:105382; Lama5.
CC DR GO: GO:0005604; C:basement membrane; IDA.
CC DR GO: GO:0030324; P:lung development; IMP.
CC DR InterPro: IPR006209; EGF_like.
CC DR InterPro: IPR000034; Laminin_B.
CC DR InterPro: IPR02049; Laminin_B.
CC DR InterPro: IPR001791; Laminin_G.
CC DR InterPro: IPR001886; LamNT.
CC DR Pfam: PF00052; laminin_B.1.
CC DR Pfam: PF00053; laminin_EGF; 18.
CC DR Pfam: PF00054; laminin_G; 2.
CC DR Pfam: PF00055; laminin_Nterm; 1.
CC DR PRINTS: PR00011; EGF_LAMININ.
CC DR ProDom: PD002082; Lam_N2; 1.
CC DR ProDom: PD003031; Laminin_B; 1.
CC DR SMART: SM00180; EGF_Lam; 20.
CC DR SMART: SM00281; Lamb; 1.
CC DR SMART: SM00282; LamG; 5.
CC DR SMART: SM00136; LamNT; 1.
CC DR PROSITE: PS00022; EGF_1; 19.
CC DR PROSITE: PS01186; EGF_2; 3.
CC DR PROSITE: PS01248; LAMININ_TYPE_EGF; 19.
CC DR PROSITE: PS00245; LAM_G_DOMAIN; 5.
CC DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC FT SIGNAL 1 40
CC FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
CC FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
CC FT DOMAIN 305 363 LAMININ EGF-LIKE 1.
CC FT DOMAIN 364 433 LAMININ EGF-LIKE 2.
CC FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
CC FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
CC FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
CC FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
CC FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
CC FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
CC FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
CC FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
CC FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
CC FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
CC FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
CC FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
CC FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
CC FT DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC FT DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
CC FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
CC FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
CC FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
CC FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
CC FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
CC FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
CC FT

SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
 RC TISSUE=Placenta;
 RX MEDLINE=97311416; PubMed=9168133;
 RA Sugaya K., Sasanuma S.-I., Nohata J., Kimura T., Fukagawa T.,
 RA Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.,
 RT "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
 RT human counterpart gene of mouse proto-oncogene Int3.";
 RL Gene 189:235-244(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Bone marrow, and Heart;
 RX MEDLINE=98360091; PubMed=9693032;
 RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
 RA Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.,
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 RT sequence of the human NOTCH4 gene.";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
 RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(IEC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3;
 CC Comment-Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=Q99466-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q99466-2; Sequence=VSP_001406;
 CC Name=3;
 CC IsoId=Q99466-3; Sequence=VSP_001407;
 CC -!- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.

CC -!- SIMILARITY: Contains 28 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D63395; BAA09708.1; ALT_FRAME.
 CC EMBL: D86566; BAAL3116.1; -
 CC EMBL: U95299; AAC32288.1; -
 CC EMBL: U89335; AAC63097.1; -
 CC EMBL: AB023961; BAB20317.1; -
 CC EMBL: AB024520; BAA88951.1; -
 CC EMBL: AB024578; BAA88952.1; -
 CC HSSP: P08709; 1BF9
 CC Genew: HGNC:7884; NOTCH4.
 CC MIM: 164951; -
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000742; EGF_2
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR001438; EGF_II.
 CC InterPro: IPR006209; EGF_like.
 CC InterPro: IPR002049; Laminin_EGF.
 CC InterPro: IPR000800; Notch.
 CC Pfam: PF00023; ank; 6
 CC Pfam: PF00008; EGF; 26.
 CC Pfam: PF00066; notch; 2.
 CC PRINTS: PRO0010; EGFBL00D.
 CC PRINTS: PRO0011; EGFLAMININ.
 CC PRINTS: PRO1452; NOTCH.
 CC SMART: SM00248; ANK; 5.
 CC SMART: SM00179; EGF_CA; 11.
 CC SMART: SM00004; NL; 3.
 CC PROSITE: PS00088; ANK_REPEAT; 5.
 CC PROSITE: PS0297; ANK_REPEAT_REGION; 1.
 CC PROSITE: PS00010; ASX_HYDROXYL; 11.
 CC PROSITE: PS00022; EGF_1; 28.
 CC PROSITE: PS01186; EGF_2; 21.
 CC PROSITE: PS01187; EGF_CA; 9.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 CC Triplet repeat expansion; Alternative splicing.
 CC SIGNAL 1 23 POTENTIAL.
 CC CHAIN 24 2003 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
 CC CHAIN 1432 2003 NOTCH EXTRACELLULAR TRUNCATION
 CC (BY SIMILARITY).
 CC CHAIN 1467 2003 NOTCH INTRACELLULAR DOMAIN
 CC (BY SIMILARITY).
 CC DOMAIN 24 1447 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 1448 1468 POTENTIAL.
 CC DOMAIN 1469 2003 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 24 63 EGF-LIKE 1.
 CC DOMAIN 64 115 EGF-LIKE 2.
 CC DOMAIN 118 155 EGF-LIKE 3.
 CC DOMAIN 156 192 EGF-LIKE 4.
 CC DOMAIN 194 232 EGF-LIKE 5.
 CC DOMAIN 234 274 EGF-LIKE 6.
 CC DOMAIN 276 312 EGF-LIKE 7.
 CC DOMAIN 314 353 EGF-LIKE 8.
 CC DOMAIN 355 391 EGF-LIKE 9.
 CC DOMAIN 392 430 EGF-LIKE 10.
 CC DOMAIN 432 473 EGF-LIKE 11.
 CC DOMAIN 475 511 EGF-LIKE 12.
 CC DOMAIN 513 549 EGF-LIKE 13.

FT	DOMAIN	551	587	EGF-LIKE 14,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	589	625	EGF-LIKE 15,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	626	659	EGF-LIKE 16,	
FT	DOMAIN	661	689	EGF-LIKE 17,	
FT	DOMAIN	691	727	EGF-LIKE 18,	
FT	DOMAIN	729	765	EGF-LIKE 19,	
FT	DOMAIN	767	803	EGF-LIKE 20,	
FT	DOMAIN	806	842	EGF-LIKE 21,	
FT	DOMAIN	844	880	EGF-LIKE 22,	
FT	DOMAIN	882	928	EGF-LIKE 23,	
FT	DOMAIN	930	966	EGF-LIKE 24,	
FT	DOMAIN	968	1004	EGF-LIKE 25,	
FT	DOMAIN	1006	1044	EGF-LIKE 26,	
FT	DOMAIN	1046	1085	EGF-LIKE 27,	
FT	DOMAIN	1087	1126	EGF-LIKE 28,	
FT	DOMAIN	1130	1171	EGF-LIKE 29,	
FT	DOMAIN	1472	1476	POLY-ARG.	
FT	REPEAT	1165	1212	LIN/NOTCH 1.	
FT	REPEAT	1213	1246	LIN/NOTCH 2.	
FT	REPEAT	1247	1286	LIN/NOTCH 3.	
FT	REPEAT	1633	1665	ANK 1.	
FT	REPEAT	1666	1698	ANK 2.	
FT	REPEAT	1700	1732	ANK 3.	
FT	REPEAT	1733	1765	ANK 4.	
FT	REPEAT	1766	1798	ANK 5.	
FT	DISULFID	28	41	BY SIMILARITY.	
FT	DISULFID	35	51	BY SIMILARITY.	
FT	DISULFID	53	62	BY SIMILARITY.	
FT	DISULFID	68	80	BY SIMILARITY.	
FT	DISULFID	74	103	BY SIMILARITY.	
FT	DISULFID	105	114	BY SIMILARITY.	
FT	DISULFID	122	133	BY SIMILARITY.	
FT	DISULFID	127	143	BY SIMILARITY.	
FT	DISULFID	145	154	BY SIMILARITY.	
FT	DISULFID	160	171	BY SIMILARITY.	
FT	DISULFID	165	180	BY SIMILARITY.	
FT	DISULFID	182	191	BY SIMILARITY.	
FT	DISULFID	198	211	BY SIMILARITY.	
FT	DISULFID	205	220	BY SIMILARITY.	
FT	DISULFID	222	231	BY SIMILARITY.	
FT	DISULFID	238	249	BY SIMILARITY.	

Alignment Scores:
Pred. No.: 0.223 Length: 2003
Score: 116.50 Matches: 116
Percent Similarity: 29.49% Conservative: 40
Best Local Similarity: 21.93% Mismatches: 191
Query Match: 3.05% Indels: 182
DB: 1 Gaps: 28

US-09-743-237-3 (1-2134) x NTC4_HUMAN (1-2003)

Qy	1	AATTCGGGGTCAAGCGNAGCTCGCGGGGGCGACAGCGGCGGGGAGGCTCTCGGGG	60
Db	740	AsnGlyGlySerCysAsnProSerProGlyGlyTyrCysThrCysProProSerHis	759
Qy	61	AGTACCCGGGATCCAGAGCTCAGCGCGCTGGAGGAGCTCGCGCTCTCGAGCCCGCG	120
Db	760	ThrGlyProGlnCysGlnThrSerThrAspTyrCysValSerAlaProCys-	776
Qy	121	AGCGCGCGCTGCAACGTGCACTTCCTCGCTGCTACCGCGCGACCGAGCCCGCG	180
Db	777	---PheAsnGlyGlyThrCysValAsnArgProGlyThrPheSerCysLeuCysAlaMet	795
Qy	181	GGGTGTTTGGCCCTGGGGCGCTGGGTCCTGCGCAAGGAGGCTCCACCGCGGGCGTCCGC	240
Db	796	GlyPheGlnGlyPro-----ArgCysGluGlyLysLeuArgPro-SerCysAl	811
Qy	241	ATGATCCAGTTGAATCAAGGTAAAGCAGGTGACTACTCAAGTAATAATCCGGAAGA	300
Db	811	aaspSerProCysArgAsnArg-----AlaThrCysGlnAspSerProGlnG	827
Qy	301	A-----GCAACTTTGCAGAAATCT	318

Db	827	yProArgCysLeuCysProThrGlyTyrThrGlyGlySerCysGlnThrLeuMetAspLe	847
Qy	319	TCITGCTCAGAAATCCCTGT-----	337
Db	847	uCysAlaGlnLysProCysProArgAsnSerHisCysLeuGlnThrGlyProSerPheH	867
Qy	338	-----TGCAAGTTCCCA-----	349
Db	867	sCysLeuCysLeuGlnGlyTrpThrGlyProLeuCysAsnLeuProLeuSerSerCysG	887
Qy	350	-----TGGTCCAGGAAGTACAGAGATGCTCTCTGT-----TCTCTTAA	390
Db	887	nLysAlaAlaLeuSerGlnGlyIleAspValSerSerLeuCysHisAsnGlyGlyLeuCy	907
Qy	391	GAAGATTCACCAATCGTGATATCCCAATTCGAAGGGGCACACAA---ATGCTATG	447
Db	907	sValAspSerGlyProSerTyrPheCysHisCysProProGlyPheGlnGlySerLeuCy	927
Qy	448	TATAGACAATCTAGAACAAAGAACTAAAA-----	478
Db	927	sGlnAspHisValAsnProCysGluSerArgProCysGlnAsnGlyAlaThrCysMetal	947
Qy	479	-----GCACCTCATTTGGTTCCTAGTATCAAGATCAAAATAATTATCT	522
Db	947	aGlnProSerGlyTyrLeuCysGlnCysAlaProGlyTyrAspGlyGlnAsnCysSerLy	967
Qy	523	ACAGTCAGAT-----GTCCTCTAAACC	543
Db	967	sGlnLeuAspAlaCysGlnSerGlnProCysHisAsnHisGlyThrCysThrProLysPr	987
Qy	544	AATGACTGCTTTAGTAGGGAGATTTTCCAGCATCAACAAATAATAATCATACACA	603
Db	987	o-----GlyGlyPheHisCysAlaCysProPro---GlyPheValGlyLe	1001
Qy	604	ACAACCTGAGGAGCCTTACATCGGTAGTCAACGGGTCTGCTTCCCTCGGAGATCAAC	663
Db	1001	uArgCysGluGlyAspValAspGluCysLeuAspGlnProCysHisProThrGlyThrAl	1021
Qy	664	TCITCCAGGACCACCAAAATAACTTTGGTGGGTACTGTGACGTCTTGCC-----AG	717
Db	1021	a-----AlaCysHisSerLeuAlaAsnAlaPheTyrCysGlnCysLeuProGlyHisTh	1039
Qy	718	TGGGAGCTTTTGC-----AACAACTGCAATTTGAATAATTTGTCACAACTT	765
Db	1039	rGlyGlnTrpCysGluValGluIleAspProCysHisSerGlnProCysPheHisGly--	1058
Qy	766	GCATCATGATATTCAACGGTTTAAAGCCATTAAGCATGCTTGGTGAATAATCCAGAAGC	825
Db	1059	-----GlyThrCysGluAlaThrAlaGly---SerProLeuG	1070
Qy	826	TTTCCAGCCAAAATTTGGAGGCGCAATTTGGGCAATGTCAAGCCCGCAGCACACAAGG	885
Db	1070	yPheileCysHisCysProLysGlyPheGluGlyProThrCysSerHisArgAlaProSe	1090
Qy	886	GTGC-----AACTCCAGAGGTACGCG-----	907
Db	1090	rCysGlyPheHisCysHisGlyGlyLeuCysLeuProSerProLysProGlyPh	1110
Qy	908	-----TGCCTGAAGAATTAC-----TGCAGTCTATGAGGC	939
Db	1110	eProProArgCysAlaCysLeuSerGlyTyrGlyGlyProAspCysLeuThrProAl	1130
Qy	940	CCAAATTTATGTGTTCTTCTATTTCAAATGCTTGGTTGCAAAATTAATGAAGAACCC	999
Db	1130	aProLysGlyCysGlyProProSerProCysLeu-----TyrAsnGlySer--	1145
Qy	1000	AGAACGAAGACACTAATAGCATGCCAACTACATGCAGACTGAGGTTTGGAGGCAG	1059
Db	1146	-----CysSerGluThrThrGlyLeuGlyGly--	1154
Qy	1060	CCATTACTCTGCCCAACGAAATTTTCAGGACTTCCAGACTTCAAGATTCAGTCAGTACGCGCC	1119

Db 1155 -----ProGly-PheArgCysSer-----C 1161
 QY 1120 TTCCTCATGCTCCTGGG---AGGTGGTGGAGGCCACATGCGCTCGCTGCTGTCPCA 1176
 Db 1161 ySProHisSerSerProGlyProArgCysGlnIlyProGlyAla---Lys-GlyCysGlu 1179
 QY 1177 GGGAGAGAGCCGAGAAAGAACACTGCTCCAACTGCTGCGCAGACGAGAT----- 1227
 Db 1180 GlyArgSerGly-----AspGlyAlaCysAspAlaGlyCys 1191
 QY 1228 ---GATCTCGAGGAATTTGG 1245
 Db 1192 SerGlyProGlyGlyAsnTrp 1198
 RESULT 13
 ID POLG_HCVBK STANDARD; PRT: 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
 OS Hepatitis C virus (Isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moomaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Van Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC CC PROTEIN C AND RNA.
 CC CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 CC DR EMBL; M58335; AAA72945.1; -
 DR PIR; A38465; GNMVTC.
 DR PDB; 1A10; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1OUV; 26-JUN-00.
 DR PDB; 8OHM; 20-APR-99.
 DR MEROPS; S29.001; -
 DR MEROPS; U39.001; -
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR004109; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RDRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RDRP; 1.
 DR ProDom; PDI86062; HCV_NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 POTENTIAL.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2041	2041	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2077	2077	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2240	2240	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2529	2529	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2788	2788	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	STRAND	1031	1035	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	HELIX	1039	1047		
FT	STRAND	1050	1050		
FT	STRAND	1059	1063		
FT	STRAND	1068	1074		
FT	TURN	1075	1076		
FT	STRAND	1077	1081		
FT	HELIX	1082	1085		
FT	TURN	1086	1087		
FT	STRAND	1090	1092		
FT	TURN	1093	1094		
FT	STRAND	1095	1097		
FT	STRAND	1101	1103		
FT	TURN	1104	1107		
FT	STRAND	1108	1112		
FT	STRAND	1120	1120		
FT	STRAND	1122	1122		
FT	TURN	1129	1133		
FT	TURN	1135	1136		
FT	STRAND	1139	1144		
FT	STRAND	1149	1157		
FT	HELIX	1158	1161		
FT	TURN	1162	1163		
FT	TURN	1165	1166		
FT	STRAND	1168	1171		
FT	TURN	1172	1174		
FT	STRAND	1175	1186		
FT	TURN	1187	1188		
FT	STRAND	1189	1197		
FT	HELIX	1198	1202		
FT	TURN	1203	1204		
FT	STRAND	1680	1688		
SQ	SEQUENCE	3010 AA; 327189 MW; F8422D5ECDFD9C CRC64;			

Alignment Scores:			
Pred. No.:	0.272	Length:	3010
Score:	116.00	Matches:	93
Percent Similarity:	34.88%	Conservative:	57
Best Local Similarity:	21.63%	Mismatches:	157
Query Match:	3.03%	Indels:	123
DB:	1	Gaps:	23

Qy	542	CCATGACTGCTTTAGTAGGAGCA	-----TTTTTGCAGCATCAACAAATTA	589
Db	1381	ProileGluAlaIleArgGlyArgHisLeuIlePheCysHisSerLysLysCys	-----TTTTTGCAGCATCAACAAATTA	1400
Qy	590	AATCTCAATACACAACTTGAGGAGCGCTTACCATCGGTATGACAGCGGTCTGCTTC	-----TTTTTGCAGCATCAACAAATTA	649
Db	1401	AspGluLeuAlaAlaLysLeuSerGlyLeu	-----GlyIleAsnAlaValAlaTyr	1417
Qy	650	CCCTGGGATCACTCTTCCAGGACCAACAAATA	-----TTTTTGCAGCATCAACAAATTA	685
Db	1418	TyArgGlyLeuAspValSerValIleProThrIleGlyAspValValAlaThr	-----TTTTTGCAGCATCAACAAATTA	1437

Qy	686	-----ACTTGGCTGGTACTGTGACTGCTTTGCCAGTGGGACTTTTGCACAACTGC	739
Db	1438	AspAlaLeuMetThrGlyTyr-----ThrGlyAspPheAspSerValIle	1452
Qy	740	AATTGTAATAATTGTTGCCAACACTTG-----CATCATGATATT	778
Db	1453	AspCysAsnThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIle	1472
Qy	779	GAACGGTTTAAAGCCATTAAAGGCATGCTTGGTAGAAATCCAGAAGCTTTCCAGCCAAAA	838
Db	1473	GluThrThrThrValProGlnAspAlaValSerArgSerGlnArg-----ArgGlyArg	1490
Qy	839	ATTGGAAGGCCCAATTGGCC-----AATGTCGAAGCCCGACGACCAACAAAGGGTGC	889
Db	1491	ThrGlyArgGlyArgArgGlyIleTyrArgPheValThrProGlyGlu-----	1506
Qy	890	AACGAGGAGGTGAGCTGCTCCGCTGAAGAT-----TACTGCGAGTGTATGAGGCCAA	943
Db	1507	-----ArgProSerGlyMetPheAspSerValLeuCysGluCysTyrAspAla---	1523
Qy	944	ATTATGTGTTCTTCTATTTCGAAATGATGCTGGTTGCAAAATTTATGAAGACCCAGAA	1003
Db	1524	-----GlyCysAlaTrpTyrGluLeuThrPro---	1532
Qy	1004	CGAAGACACTAATGAGCATGCCAACTACATGCAGACTGGAGGTTTGGAA-----	1054
Db	1533	AlaGluThrSerValArgLeuArgAlaTyrLeuAsnThrProGlyLeuProValCysGln	1552
Qy	1055	-----GGCAGCCATTAC	1066
Db	1553	AspHisLeuGluPheTrpGluSerValPheThrGlyLeuThrHisIleAspAlaHisPhe	1572
Qy	1067	CTGCCACCAACGAAATTTTTCAGGA-----CTTCCAAAGATTACGTC-----AC	1108
Db	1573	LeuSerGlnThrLysGlnAlaGlyAspAsnProTyrLeu-ValAlaTyrGlnAlaTh	1592
Qy	1109	GATAGGCGCTTCTCATCATCTCTCTGGAGGTGG-----TGGAGGCCACATGC	1159
Db	1592	rValCysAlaArgAlaGlnAlaProProSerTrpAspGlnMetTrpLys-----	1609
Qy	1160	GCCTGCTGCTTCTCAGGAGGAGAGAGCCGAGAACACACTGCTCCAGTCCCTGGCA	1219
Db	1610	-----CysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrAr	1627
Qy	1220	GAGCAGATGATCTCTGGAGGAATTTGGAAGGTGCTTATCACAGATTCTCCACACTGACT--	1277
Db	1627	gLeu-----GlyAlaValGlnAsnGluValThrLeuThrHi	1639
Qy	1278	-----TTAAATCTAAGGGATTGAAA	1297
Db	1639	sProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValValThrSerTh	1659
Qy	1298	ATGAGTAGAGTATAAAGTGTGAATGCATGTTGCTTGTCTAGTCTAGAAATCTCTAG	1357
Db	1659	rTrpValLeuValGlyGlyValLeuAlaLeuAlaTyrCysLeuThrThrGlySe	1679
Qy	1358	TTTAGAAAGGATGTTTAGGGGAACATGAGGCTGCTCTGAGCAACAAACAGGCTCCCT	1417
Db	1679	rValValIleVal-----GlyArgIleLeuSerGlyArgProAlaIle	1694
Qy	1418	GCATCCTCGGCCAGGAGGATTTACTGACAGCTCTCT-----GAGATGTGGCAACCCA	1471
Db	1694	eValProAspArgGluLeuLeuTyrGlnGluPheAspGluMetGluGluCysAlaSerHi	1714
Qy	1472	TGCCCCCTTTCTGAGGAGGTGCTGCTGAGCATTTGTTGCTGCCCCAGAGGAGAGA	1531
Db	1714	sLeuProTyrIleGln-----GlyMetGlnLeuAlaGluGlnPheLysGlnLysAl	1732
Qy	1532	GCTTGGGTTCCCATAGTCTGAGGAGGTGTCTGAGGCGCGGCGGAGGAGGAGCCGCC	1591
Db	1732	aLeuGlyLeu-----LeuGlnThrAlaThrLysGlnAlaGluAla-----AlaAlaPr	1748

Pfam:	PF01542;	HCV_core; 1.
DR	Pfam:	PF01539; HCV_env; 1.
DR	Pfam:	PF01560; HCV_NS1; 1.
DR	Pfam:	PF01538; HCV_NS2; 1.
DR	Pfam:	PF02907; HCV_NS3; 1.
DR	Pfam:	PF01006; HCV_NS4a; 1.
DR	Pfam:	PF01001; HCV_NS4b; 1.
DR	Pfam:	PF01506; HCV_NS5a; 1.
DR	Pfam:	PF00271; helicase_C; 1.
DR	Pfam:	PF00998; Viral_RdRP; 1.
DR	ProDom:	PD186062; HCV_NS1; 1.
DR	SMART:	SM00487; DEXdc; 1.
KW	Polyprotein:	Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW	Core protein:	Coat protein; Envelope protein; Helicase; ATP-binding;
KW	Transmembrane:	Nonstructural protein; Hydrolase; Serine protease;
3D-structure:		
FT	INIT_MET	1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.
FT	CHAIN	1 115 CAPSID PROTEIN C (POTENTIAL).
FT	CHAIN	116 191 MATRIX PROTEIN E (POTENTIAL).
FT	CHAIN	192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT	CHAIN	384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT	CHAIN	730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT	CHAIN	1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT	CHAIN	1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT	CHAIN	1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT	CHAIN	2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT	TRANSMEM	347 369 POTENTIAL.
FT	ACT_SITE	1083 1093 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	NP_BIND	1230 1237 ATP (POTENTIAL).
FT	SITE	1316 1319 DECH BOX.
FT	CARBOHYD	196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SEQ	SEQUENCE	3010 AA; 326573 MW; 94ALC77435D642BB CRC64;

Alignment Scores:
 Pred. No.: 0.272 Length: 3010
 Score: 116.00 Matches: 93
 Percent Similarity: 35.12% Conservative: 58
 Best Local Similarity: 21.63% Mismatches: 156
 Query Match: 3.03% Indels: 123
 DB: 1 Gaps: 24

US-09-743-237-3 (1-2134) x POLG_HCVJTT (1-3010)
Qy 542 CCAATGACTCTTATTAGTAGGGAGA-----TTTTTCAGCATCATCAAAATA 589
: :
Db 1381 ProLeuAlaIleLysGlyArgHisLeullehecYhisSerLysLysLysCys 1400
: :
Qy 590 AATCTCATTTACACAACAACCTTGAGGAGCGCTTACCATCGGTAGTCACACGGCTGCTGTTTC 649
::: :::: :
Db 1401 AspGluLeuAlaAlaLysLysSerGlyLeu-----GlyIleAsnAlaValAlaTyr 1417

QY 650 CCCTGGGA-----TCAACTCTTCCAGGACCACCAAAAATAACT----- 688
 Db 1418 TyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValIleValAlaThr 1437
 QY 689 -----TTGGCTGGTACTGTGACTGCTTTGCCAGTGGGACTTTTTCACAACTGC 739
 Db 1438 AspAlaLeuMetThrGlyTyr-----ThrGlyAspPheAspSerValIle 1452
 QY 740 AATTGTAATAATTCTGCAACAACCTTG-----CATCATGATATT 778
 Db 1453 AspCysAsnThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIle 1472
 QY 779 GAACGGTTTAAGCCATTAAAGCATGTCTTGGTAGAAATCCAGAGCTTTTCCAGCCAAA 838
 Db 1473 GluThrThrThrValProGlnAspAlaValSerGlnArg-----ArgGlyArg 1490
 QY 839 ATTGGGAAGGCCAATTGGGC-----AATGTCAGCCCCAGCACAACAAAGGTGC 889
 Db 1491 ThrGlyArgGlyArgGlyGlyIleTyrArgPheValThrProGlyGlu----- 1506
 QY 890 AACTGCAGAGGTGAGGCTGCTGAAGAAT-----TACTCGAGTGTATGAGGCCAA 943
 Db 1507 -----ArgProSerGlyMetPheAspSerValLeuCysGluCysTyrAspAla--- 1523
 QY 944 ATTATGTCTTCTTATTGCAATGTCATGTTGCAAAATATGAGAAAGCCAGAA 1003
 Db 1524 -----GlyCysAlaTyrTyrGluLeuThrPro--- 1532
 QY 1004 CGAAGACACTAATGACATGCCAACTACATGACAGCTGAGGTTTGGAA----- 1054
 Db 1533 AlaGluThrThrValArgLeuArgAlaTyrLeuAsnThrProGlyLeuProValCysGln 1552
 QY 1055 -----GGCAGCCATTAC 1066
 Db 1553 AspHisLeuGluPheTrpGluSerValPheThrGlyLeuThrHisIleAspAlaHisPhe 1572
 QY 1067 CTGCACCAACGAAATTTTCAGGA-----CTTCCAAGATTCAGTC-----AC 1108
 Db 1573 LeuSerGlnThrLysGlnAlaGlyAspAsnPheProTyrLeu-ValAlaTyrGlnAlaTh 1592
 QY 1109 GATAGCGCGCTTCTCATGATCTCTGGGAGTGG-----TGGAGGCCACATGC 1159
 Db 1592 rValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLys----- 1609
 QY 1160 GCCTCGCTCTGCTCAGGAGAGAGAGCCGAGAAAGAACACTGCTCCTCAAGTGGCTGGCA 1219
 Db 1610 -----CysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrAr 1627
 QY 1220 GAGCAGATGATCTGGAGGAATTTGGAGGTGCTTATCAGAGATCTCCACACTGAGT-- 1277
 Db 1627 gLeu-----GlyAlaValGlnAsnGluIleThrLeuThrHi 1639
 QY 1278 -----TTAAATCTAAGGATTGAAA 1297
 Db 1639 sProIleThrLysPheIleMetAlaCysMetSerAlaAspLeuGluValValThrSerTh 1659
 QY 1298 ATGGAGTAGATATAAAGTGTGAATGCTGTTGTTTGTCTAGTCTAGAAATCTCTAG 1357
 Db 1659 rTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuThrThrGlySe 1679
 QY 1358 TTTAGAAGGATGTTTAGGGGAACATGAGGCTGCTGTCAGCAACACAGGCTCCCT 1417
 Db 1679 rValValIleVal-----GlyArgIleIleLeuSerGlyArgProAlaVa 1694
 QY 1418 GCATCCCTGGCCCGGAGGAGTTTACTCAGACTCTCT-----GAAGATGGGCAACCCA 1471
 Db 1694 lValProAspArgGluValLeuTyrArgGluPheAspGluMetGluCysAlaSerHi 1714
 QY 1472 TGCCCTCTTTTCTGAGGAGGTGCATGGCTGAGCATTGTTTGTCTGGCCCGCAGGAGAGA 1531
 Db 1714 sLeuProTyrIleGluGln-----GlyMetGlnLeuAlaGluGlnPheLysGlnLysAl 1732
 QY 1532 GCTTGGGTTCCTAGTCTGGGAGAGTGTCTGCAAGCGCGCGGAGGCGAGCAGGCC 1591

Db 1732 aLeuGlyLeu-----LeuGlnThrAlaThrLysGlnAlaGluAla----AlaAlaPr 1748
 QY 1592 TGGCGAGAGCTCACTCTGGTGGACTCTT 1619
 Db 1748 oValValGluSerArgTrpArgAlaLeu 1757
 RESULT 15
 ITB2_HUMAN
 ID ITB2_HUMAN STANDARD; PRT; 769 AA.
 AC P05107; Q16418;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CR3/p150,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit).
 DE ITGB2 OR CD18.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87131080; PubMed=3028646;
 RA Kishimoto T.K., O'Connor K., Lee A., Roberts T.M., Springer T.A.;
 RT "Cloning of the beta subunit of the leukocyte adhesion proteins:
 RT homology to an extracellular matrix receptor defines a novel
 RT supergene family.";
 RL Cell 48:681-690(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92077153; PubMed=1683838;
 RA Weitzman J.B., Wells C.E., Wright A.H., Clark P.A., Law S.K.A.;
 RT "The gene organisation of the human beta 2 integrin subunit (CD18).";
 RL FEBS Lett. 294:97-103(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [4]
 RP SEQUENCE OF 9-769 FROM N.A., AND PARTIAL SEQUENCE.
 RN TISSUE-Spleen;
 RX MEDLINE=87246525; PubMed=2954816;
 RA Law S.K.A., Gagnon J., Hildreth J.E., Wells C.E., Willis A.C.,
 RA Wong A.J.;
 RT "The primary structure of the beta-subunit of the cell surface
 RT adhesion glycoproteins LFA-1, CR3 and p150,95 and its relationship to
 RT the fibronectin receptor.";
 RL EMBO J. 6:915-919(1987).
 RN [5]
 RP SEQUENCE OF 347-355 FROM N.A., AND VARIANTS LAD1 SER-351 AND TRP-586.
 RX MEDLINE=92147694; PubMed=1346613;
 RA Nelson C., Rabb H., Arnaut M.A.;
 RT "Genetic cause of leukocyte adhesion molecule deficiency. Abnormal
 RT splicing and a missense mutation in a conserved region of CD18 impair
 RT cell surface expression of beta 2 integrins.";
 RL J. Biol. Chem. 267:3351-3357(1992).
 RN [6]
 RP VARIANTS LAD1 PRO-149 AND ARG-169.

RX MEDLINE-90293696; PubMed-1694220;
RA Wardlaw A.J., Hibbs M.L., Stacker S.A., Springer T.A.;
RT "Distinct mutations in two patients with leukocyte adhesion
RL deficiency and their functional correlates";
RJ J. Exp. Med. 172:335-345(1990).
RN [7]
RX VARIANT LAD1 ARG-169.
RP MEDLINE-92324303; PubMed-1352501;
RA Corbi A., Vara A., Ursa A., Rodriguez M.C.G., Fontan G.,
RJ Sanchez-Madrid F.;
RT "Molecular basis for a severe case of leukocyte adhesion deficiency.";
RL Eur. J. Immunol. 22:1877-1881(1992).
RN [8]
RX VARIANTS LAD1 THR-196 AND CYS-593.
RP MEDLINE-90187099; PubMed-1968911;
RA Arnout M.A., Dana N., Gupta S.K., Tenen D.G., Fathallah D.M.;
RT "Point mutations impairing cell surface expression of the common beta
RL subunit (CD18) in a patient with leukocyte adhesion molecule
RJ (Leu-CAM) deficiency.";
RN J. Clin. Invest. 85:977-981(1990).
RX VARIANT LAD1 LEU-178.
RP MEDLINE-92184805; PubMed-1347532;
RA Back L.L., Kwok W.W., Hickstein D.D.;
RT "Identification of two molecular defects in a child with leukocyte
RJ adherence deficiency.";
RL J. Biol. Chem. 267:5482-5487(1992).
RN [10]
RX VARIANT LAD1 ASN-128.
RP MEDLINE-92272746; PubMed-1590804;
RA Katsura S., Kishi F., Tsukahara M., Nunoi H., Matsuda I.,
RJ Kobayashi K., Kajii T.;
RT "Leukocyte adhesion deficiency: identification of novel mutations in
RL two Japanese patients with a severe form";
RJ Biochem. Biophys. Res. Commun. 184:1460-1467(1992).
RN [11]
RX VARIANT LAD1 SER-284.
RP MEDLINE-93312347; PubMed-7686755;
RA Back L.A., Kerkerling M., Baker D., Bauer T.R., Embree L.J.,
RJ Hickstein D.D.;
RT "A point mutation associated with leukocyte adhesion deficiency type
RJ 1 of moderate severity";
RL Biochem. Biophys. Res. Commun. 193:912-918(1993).
RN [12]
RX VARIANTS LAD1 PRO-138 AND ARG-273.
RP MEDLINE-99102253; PubMed-9884339;
RA Hogg N., Stewart M.P., Scarth S.L., Newton R., Shaw J.M., Law S.K.A.,
RJ Klein N.;
RT "A novel leukocyte adhesion deficiency caused by expressed but
RL nonfunctional beta2 integrins Mac-1 and LFA-1";
RJ J. Clin. Invest. 103:97-106(1999).
CC -1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
CC ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE
CC RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT
CC AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE
CC SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2
CC RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN
CC ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-
CC D/BETA 2 IS A RECEPTOR FOR ICAM3 AND VCAM1.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2
CC ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: DEFECTS IN ITGB2 ARE THE CAUSE OF LEUKOCYTE ADHESION
CC DEFICIENCY, TYPE I (LAD1). LAD1 PATIENTS HAVE RECURRENT BACTERIAL
CC INFECTIONS AND THEIR LEUKOCYTES ARE DEFICIENT IN A WIDE RANGE OF
CC ADHESION-DEPENDENT FUNCTIONS.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC -1- SIMILARITY: Contains 2 VWFA-like domains.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD18 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd18.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00057; CAA68266.1; -
DR EMBL; X64072; CAA45427.1; -
DR EMBL; X64073; CAA45427.1; JOINED.
DR EMBL; X64074; CAA45427.1; JOINED.
DR EMBL; X64075; CAA45427.1; JOINED.
DR EMBL; X64076; CAA45427.1; JOINED.
DR EMBL; X64077; CAA45427.1; JOINED.
DR EMBL; X64078; CAA45427.1; JOINED.
DR EMBL; X64079; CAA45427.1; JOINED.
DR EMBL; X64080; CAA45427.1; JOINED.
DR EMBL; X64081; CAA45427.1; JOINED.
DR EMBL; X64082; CAA45427.1; JOINED.
DR EMBL; X64083; CAA45427.1; JOINED.
DR EMBL; X63924; CAA45427.1; JOINED.
DR EMBL; X63925; CAA45427.1; JOINED.
DR EMBL; X63926; CAA45427.1; JOINED.
DR EMBL; AL163300; CAB90553.1; -
DR EMBL; M15395; AA59490.1; -
DR EMBL; S81234; AAB21404.1; -
DR PIR; A25967; IJHULM.
DR PDB; 1JX3; 19-SEP-01.
DR PDB; 1L3Y; 01-APR-02.
DR Genew; HGNC:6155; ITGB2.
DR MIM; 600065; -
DR MIM; 116920; -
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0004895; F:cell adhesion receptor activity; TAS.
DR GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. .; TAS.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; TAS.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002369; Integrin_B.
DR InterPro; IPR001169; Integrin_beta_C.
DR InterPro; IPR003659; Plexin-like.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00362; Integrin_B; 1.
DR PRINTS; PRO1186; INTEGRINB.
DR ProDom; PD001811; Integrin_B; 1.
DR SMART; SM00187; INB; 1.
DR SMART; SM00423; PSI; 1.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00243; INTEGRIN_BETA; 3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Signal; Disease mutation; Pyrrolidone carboxylic acid;
KW 3D-structure.
FT SIGNAL 1 22
FT CHAIN 23 769 INTEGRIN BETA-2.
FT DOMAIN 701 723 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 724 769 POTENTIAL.
FT DOMAIN 124 363 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 449 617 VWFA-LIKE.
FT DOMAIN 449 496 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 449 496 I.
FT REPEAT 497 540 II.
FT REPEAT 541 581 III.
FT REPEAT 582 617 IV.
Alignment Scores:
Pred. No.: 0.21 Length: 769
Score: 115.50 Matches: 37
Percent Similarity: 34.16% Conservative: 59
Best Local Similarity: 21.00% Mismatches: 101
Query Match: 3.02% Indels: 84
DB: 1 Gaps: 16
US-09-743-237-3 (1-2134) x ITB2_HUMAN (1-769)

```
QY 304 AACTTTGCAGATCTTCTGCTCAGGAATCCTGTTGCAAGTTCCATGGTCCCGAGAACT 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 AsnGlnProArgGlyAspCysAspGlyValGlnIleAsnValProIle-ThrPheGlnVa 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 364 AGAGATGCCCTCCCTGCTCTCTCTTAAGAAGATTCC----- 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 lLysValThrAlaThrGluCysIleGlnGlnSerPheValIleArgAlaLeuGlyPh 433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 401 -AACCCAAATGCTGATATGCCAAATGAAGAGGGGCACACAAATGCTATATAGACAATTC 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 eThrAspIleValThrValGlnValLeuProGlnCysGlyCysArgAspGlnSe 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 TAGAACAGAGAACTAAAGACACTCCATTTGGTTCTCAGTATCAAGATCAAAATAATTA 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 453 rArgAspArgSerLeu-----CysHis-----GlyLysGlyPh 464
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 TCTACAGTCAGATGCCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTGCAGCATC 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 eLeuGluCysGlyIleCysArgCysAspThrGlyTyrIleGlyLysAsnCysGluCysGl 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 580 AACAAATTAATCTCATTACACAACTTGGAGGAGCC----- 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 nThrGln---GlyArgSerSerGlnGluLeuGluGlySerCysArgLysAspAsnSe 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 620 -----TTACCATCGGTAGTCAACGGGTCTGCTTTCCCTCGGGATCAAC 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 rIleIleCysSerGlyLeuGlyAspCysValCysGlyGlnCysLeuCysHisThrSerAs 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 664 TCTTCCAGGACCAACAAATAACTTTGGCTGGGTACTGTGACTGCTTGGC----- 715
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 523 pValProGly-----LysLeuIleTyrGlyGlnTyrCysGluCysAspThrIleAsnC 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 716 -----AGTGGGACTTTTGCACAAC-----TGCAATTGTAATAA 750
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 sGluArgTyrAsnGlyGlnValCysGlyGlyProGlyArgGlyLeuCysPheCysGlyLy 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 751 TTGTTGCAACAACCTGCATCATGATATTGAACGGTTT-----AAAGCCAT 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 sCys-----ArgCysHisProGlyPheGluGlySerAlaCysGlnCysGluArgThrTh 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 796 TAAGCATGCTTGGTAGAATCCAGAGCTTTCCAGCCAAAATTTGGGAAGGCCAATT 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 579 rGluGlyCysLeu-----AsnProArgArgValGluCysSer---GlyArgGlyArg-- 595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 GGGCAATCTCAAGCCCGCCAGCACAAAGGGTGCAGCTCAGGAGGTGAGCTGCCTCAA 915
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 -----CysArgCys----- 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 916 GAATTACTGCGAGTGTATAGAGCCCAATATG-----TG 951
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 -AsnValCysGluCysHisSerGlyTyrGlnLeuProLeuCysGlnGluCysProGlyCy 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 952 TTCTTCTATTTC-----AATGCTTGGTTGCAAAATTTATGAAGAAGCCC 999
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 sProSerProCysGlyLysTyrIleSerCysAlaGluCysLeuLysPheGluLysGlyPr 638
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1000 A 1000
   |
Db 638 O 638
```

Search completed: July 24, 2003, 14:18:55
Job time : 104.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:49:34 ; Search time 69.0197 Seconds

(without alignments)

5946.827 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcggggtcaagcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p_model -DEV=xlp
-Q/cgn2_1/USPTO-spool/US09743237/runat_24072003_113105_17684/app_query.fasta_1.4750
-DB=pir_76 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237.qcgn_1_1_228 @runat_24072003_113105_17684 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	437	11.4	429	2 T23152	hypothetical prote
2	376	9.8	571	2 B84585	hypothetical prote
3	358	9.4	603	2 T08955	hypothetical prote
4	248	6.5	658	2 F71410	hypothetical prote
5	123	3.2	1700	2 S08167	Balbani ring 3 pr
6	121	3.2	1737	2 T00209	MEGF8 protein - hu
7	119.5	3.1	1001	2 T00532	probable cadmium-t
8	119.5	3.1	1172	2 F84572	probable cadmium-t
9	119.5	3.1	3010	1 GNVVCJ	genome polyprotein
10	119	3.1	1895	2 T15881	hypothetical prote
11	118.5	3.1	1801	1 M8RTS	laminin beta-2 cha
12	118.5	3.1	5376	2 T42215	zonadhesin - mouse
13	118	3.1	2406	2 A54148	odz protein - frui
14	118	3.1	2515	2 S47008	tenascin-like prot

15	117.5	3.1	3507	2 T34513	hypothetical prote
16	117	3.1	497	2 T27827	hypothetical prote
17	117	3.1	3011	1 GNVVC3	genome polyprotein
18	117	3.1	3635	2 T10053	laminin alpha 5 ch
19	116	3.0	3010	1 GNVVTC	genome polyprotein
20	116	3.0	3010	1 A45573	genome polyprotein
21	115.5	3.0	769	1 IJHULM	leukocyte adhesion
22	115.5	3.0	1766	2 A42125	trophozoite cystei
23	115.5	3.0	2195	2 T34264	hypothetical prote
24	114.5	3.0	778	2 T38487	tastin - human
25	114.5	3.0	3759	2 A35085	trithorax protein
26	113.5	3.0	328	2 G89152	protein C24B5.5 [i
27	113	3.0	1820	2 A55494	latent transformin
28	112.5	2.9	574	2 B88465	protein B0244.8 [i
29	112.5	2.9	3133	2 S52093	hemocytin - silkw
30	112	2.9	1599	2 T16210	hypothetical prote
31	112	2.9	3010	1 GNVVTV	genome polyprotein
32	111	2.9	798	2 A40526	integrin beta-7 ch
33	110	2.9	1819	2 A71928	cag island protein
34	109.5	2.9	573	2 A33533	cell surface glyco
35	109.5	2.9	3010	1 S18030	genome polyprotein
36	109	2.9	2180	2 T29764	hypothetical prote
37	109	2.9	2499	1 A30788	mannose 6-phosphat
38	108.5	2.8	1548	2 S34583	serine proteinase
39	108.5	2.8	2225	2 T26063	hypothetical prote
40	108	2.8	1607	2 T02837	long chain fatty a
41	108	2.8	1840	2 T30250	GPI protein - mous
42	108	2.9	5376	2 T42215	zonadhesin - mouse
43	107	2.8	559	2 A57474	extracellular matr
44	106.5	2.8	1786	1 MMHUB1	laminin beta-1 cha
45	106.5	2.8	1797	2 A55677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T23152

hypothetical protein JC8.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23152

R:Lightning, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19699

A:Accession: T23152

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-429 <WIL>

A:Cross-references: EMBL:282274; PIDN:CAB05228.1; GSPDB:GN00022; CESP:JC8.6b

A:Experimental source: clone JC8

C:Genetics:

A:Gene: CESP:JC8.6b

A:Map position: 4

A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Alignment Scores:

Pred. No.:	4.83e-28	Length:	429
Score:	437.00	Matches:	92
Percent Similarity:	56.95%	Conservative:	35
Best Local Similarity:	41.26%	Mismatches:	62
Query Match:	11.43%	Indels:	34
DB:	2	Gaps:	8

US-09-743-237-3 (1-2134) x T23152 (1-429)

QY	698	TACTGTGACTGCTTTGCCAGTGGGACATTTTGCACAACTGCAATTTGAATAATTTGTCG 757
Db	182	TyrCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysSHLS 201
QY	758	ACAACTTCATCATGATTTGAAGCGTTTAAGGCATTAAGGCATTTGCTTGGTGAAGAT 817
Db	202	AsnAsnIleGluTyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 221

```

Qy 818 CCAGAGCTTCCAGCCCAAAATTTGGG-----AAGGCCCAATTTGGCAATGTCTCAAGCCC 871
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 ProAsnAlaPheLysProLysIleGlyLeuAlaArgGlyGlyIleThrAspIleGluArg 241
Qy 872 CAGCACAAAGAGGTGCAACTGCAGGAGTCAGGCTGCCTGAGAAATTTACTGCGAGTGC 931
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrCysGluCys 261
Qy 932 TATGAGGCCAAATTTGTTCTTCTATTGCAAAATGTCATTGTTGTCACAAATTT----- 985
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
262 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 281
Qy 986 -----TATGAAGAAAGC-----CCAGAACGAAAGACACTA 1015
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 ThrTyrArgMetThrArgTyrLysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeu 301
Qy 1016 ATGAGCATGCCAAACTACATCCAGACT-----GGAGGTTTG--- 1051
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 MetSerLeuThrAsnAlaSerSerThrAlaThrProAspSerGlyProGlySerValVal 321
Qy 1052 -----GAAGGCAGCCATTACTCGCCCAACAGAAATTTTCAGGACTTCCAGATTC 1102
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 ThrAspGluHisGlyAspAspTyr---GluAspMetLeuLeuSerHisLysProLysVal 340
Qy 1103 AGTCAGCATAGCCGG-----CCTTCCTCAGCATCCTCGGAGGTGGTGAGGCCACACA 1156
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 GluMetAspProArgArgPheProTyrTyrMetThrAspGluValValGluAlaAla 360
Qy 1157 TGCCGCTCGCTGCTGCTCAGGAGAGAGAGCC-----GAGAAAGAA 1198
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 ThrMetCysMetValAlaGlnAlaGluAlaLeuAsnTyrGluLysValGlnThrGlu 380
Qy 1199 CACTGCTCCAAAGTCCCTGCGCAGACAGATGATCCTGGAGGAATTTGGAAGGTGCTTATCA 1258
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 AspGluLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGlu 400
Qy 1259 CAGATTCTC 1267
Db |||||:|||||:
401 GlnMetIle 403

RESULT 2
B84585
hypothetical protein At2g20110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84585
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84585
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <STO>
A:Cross-references: GB:AE002093; NID:94580462; PIDN:AAD24386.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g20110
A:Map position: 2

Alignment Scores:
Pred. No.: 5,71e-23 Length: 571
Score: 376.00 Matches: 95
Percent Similarity: 44.27% Conservative: 48
Best Local Similarity: 29.41% Mismatches: 90
Query Match: 9.83% Indels: 2
Gaps: 10
DB: 2

US-09-743-237-3 (1-2134) x B84585 (1-571)
Qy 527 TCAGATGTCCTTAACCAATCACTGCTTTAGTAGGAGATTTTTCAGGAGCATCAACAAAA 586
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3
T08955

hypothetical protein F19B15.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08955
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro

```

Db 51 AlaSerIleProSerProIleValThrValThrArgProIleIleThrSerGlnAlaPro 70
Qy 587 TTAAATCTCATTTACA-----CAACAACCTTGAGGAGGCTTA----- 622
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 ProThrValAlaThrProIleProProProGlnSerGlnGlyIleIleLeuHisVal 90
Qy 623 -----CCATCGGTAGTCAACGGGTCTGCTTTCCCTCCGATCAACT 664
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 ProIleArgHisProArgProGluSerProAsnSerMetProArgProAlaGlyGluThr 110
665 CTTCAGAGCACCACAAA-----ATRAACTTTGGCT 694
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 ArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHisSerArgCysLeuLys 130
695 GGGTACTGTGACTGCTTTGCCAGTGGGACATTTTGCACAACTCAATTTGTAATAATGT 754
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 LeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCysAsnCysValAsnCys 150
755 TGCAACAACCTTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGCTCTTGGTAGA 814
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 PheAsnAsnValGluAsnGluProAlaArgGlnAlaValGluSerThrLeuGluArg 170
815 AATCCAGAAGCTTTCCAGCCCAAAATTT-----GGGAAG 847
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 AsnProAsnAlaPheArgProLysIleAlaAlaSerProHisGlyGlyArgAspAsnArg 190
848 GGCCAAATTTGGCAATGTC-----AAGCCCGACACAAAGGGTGCACACTGCAGGAGG 901
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 GluGluValGlyAspValValMetLeuAlaArgHisAsnLysGlyCysHisCysLysLys 210
902 TCAGGCTGCCTGGAAGATTACTGCGAGTGCATGAGGCCCAAAATTTATGTTCTTCATT 961
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 SerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIleLeuCysSerGluAsn 230
962 TGCAAAATGCTTGTCTCCAAAATTTATCAAGAAATTCAGAACCCAGAACGACACTAATGAGC 1021
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 CysLysCysLeuAspCysLysAsnPheGluGlySerGluValArgGlnSerLeuPheHis 250
1022 ATGCCAAAC-----TACATGCAG-----ACTGGAGGT 1048
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 GlyGluHisSerHisAsnLeuAlaTyrLeuGlnHisAlaAsnAlaAlaIleThrGlyAla 270
1049 TTGGAAGCGAGCCATTTACCTG-----CCACCAAGAAATTTTCAGGA----- 1090
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 IleGlySerSerGlyPheAlaSerAlaProProLysArgArgLysGlyGlnGluIle 290
1091 -----CTTCCAAGATTCAGTCACGAT 1111
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 PhePheAsnGlnGlyThrLysAspSerSerThrHisArgLeuGlyGlnAlaAsnGly 310
1112 AGCGCGCTTCTCTCA-----TGCACTCTCGGAGGTGGTGGAGGCCACA 1156
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 ArgThrThrSerSerGlnThrGlySerArgAlaGlyGlyAsnAlaSerLeuGlyProSer 330
1127 -----TGCATCTCTCGGAGGTGGTGGAGGCCACA 1156
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 LysValValTyrLysSerLeuLeuAlaAsnIleIleLysProMetAspValLysAlaLeu 350
1157 TGCCTCTGCTGCTGCTGCTCAGGAGAGAGCCCGAGAAAGAACACTGCTCCAACTGCTGT 1216
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 CysSerValLeuValAlaValAlaGlyGluAlaAlaLysThrLeuThrGluLysArgLeu 370
1217 GCAGAGCAG 1225
Db |||||:|||||:
371 AlaAsnGln 373

RESULT 3
T08955
hypothetical protein F19B15.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08955
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro

```


submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16519
 A:Accession: T08955
 A:Molecule type: DNA
 A:Residues: 1-603 <BEV>
 A:Cross-references: EMBL:AL078470; GSPDB:GN00062; ATSP:F19B15.30
 A:Experimental source: cultivar Columbia; BAC clone F19B15
 C:Genetics:
 A:Gene: ATSP:F19B15.30
 A:Map position: 4
 A:Introns: 107/2; 144/2; 203/3; 321/3; 356/2; 388/1; 522/1

Alignment Scores:
 Pred. No.: 1.79e-21 Length: 603
 Score: 358.00 Matches: 93
 Percent Similarity: 44.5% Conservativeness: 41
 Best Local Similarity: 30.90% Mismatches: 76
 Query Match: 9.36% Indels: 91
 DB: 2 Gaps: 12

US-09-743-237-3 (1-2134) x T08955 (1-603)

QY 521 CTACAGTGCAGTGTCCCT-----AAACCA---ATGACTGCTTTAGTAGGAGATT 568
 Db 58 IISerSerSerValProSerThrIleArgProGlyMetThrIleAlaIleGlyGlnVal 77
 QY 569 -----TTGCCAGCATCAACAAATTAATCTCATTACACAACTT 610
 Db 78 ThrGlnValArgProThrLeuProMetAlaThrMetSerAsnProSerGlnSer 97
 QY 611 GAGGAGCGCTTACCATCGGTAGTCAACGGGTCTGCTTCCCTCGGATCACTCTTCCA 670
 Db 98 Gln-----IleValAsnAla-----ProIleArgHisProIlePro 109
 QY 671 GGACCACCAAAATA----- 685
 Db 110 GluSerProLysAlaArgGlyProArgProAsnValGluArgAspGlyThrProGln 129
 QY 686 -----ACTTTGGCTGGTACTGTGACTGCTTT 712
 Db 130 LysLysLysGlnCysAsnCysLysHisSerArgCysLysLysLeuTyrcysGluCysPhe 149
 QY 713 GCCATGGGAGCTTTTGCACAACTGCAATGTAATAATTTGTGCAACAACTGTCATCAT 772
 Db 150 AlaSerGlyThrTyrcysaspGlyCysAsnCysValAsnCysPheAsnValAspAsn 169
 QY 773 GATATTGAACGCTTTAAAGCCATTAAAGCATCTTGGTAGAAATCCAGACCTTCCAG 832
 Db 170 GluProAlaArgArgGluAlaValGluAlaThrLeuGluArgAsnProPheAlaPheArg 189
 QY 833 CCAAAATTTGGAAGGCCAATTTGGCAAT----- 862
 Db 190 ProLysIleAlaSerSerProHisGlyArgAspLysArgGluAspIleGlyGluVal 209
 QY 863 ----GTCAAGCCCGACACAAAGGGTCAACTCGAGGAGTTCAGGCTGCTGAAGAAT 919
 Db 210 ValLeuLeuGlyLysHisAsnLysGlyCysHisLysSerGlyCysLeuLysLys 229
 QY 920 TACTCGAGTGTATGAGGCCCAATTTATGCTTCTTCTATTGCAAAATGCTTGGTTC 979
 Db 230 TyrCysGluCysPheGlnAlaAsnIleLeuCysSerGluAsnCysLysCysLeuAspCys 249
 QY 980 AAAAAATTGAAGAAGCCGACAGCAAGACACTAATGAGCATGCCAAAC----- 1030
 Db 250 LysAsnPheGluGlySerGluGluArgGlnAlaLeuPheHisGlyGluHisSerAsnHis 269
 QY 1031 -----TACATCAG-----ACTGGAGGTTTGAAGCGCAGCCAT 1063
 Db 270 MetAlaTyrluGlnGlnAlaAlaAsnAlaAlaIleThrGlyAlaValGlySerSerGly 289
 QY 1064 TACCTGCCA-----CCACGAAATTTTCAGCA----- 1090
 Db 290 PheAlaProSerProAlaLysArgArgLysGlnGluIleLeuPheAsnGlnAla 309

QY 1091 CTTCCAAGATTTCAGTCAGGATAGCGGCTTCTCTCATGCTCTCTGGAGGTGCTGGAG 1150
 Db 310 Ile-LysAspSerSerArgLeuSerHisPheProGlnValAsnAsnGly----- 325
 QY 1151 GCCACATGGCTGCTGCTGCTCAGGAGGAAGCGGCGAGAAACACATGCTCCCAAG 1210
 Db 326 -----ArgThrGlyGlyProThrSerGlyThrSerProSe 337
 QY 1211 T 1211
 Db 337 r 337

RESULT 4

F71410
 hypothetical protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 A:Variety: columbia
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
 C:Accession: F71410

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
 avanagh, T.; Hempel, S.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
 erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
 C.: Chalwatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
 A:Reference number: A71400; MUID:98121113; PMID:9461215
 A:Accession: F71410
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-658 <BEV>
 A:Cross-references: GB:297337; NID:g2244829; PID:e326824; PID:g2244834
 C:Genetics:
 A:Map position: 4COP9-4G3845

Alignment Scores:

Pred. No.: 2.49e-12 Length: 658
 Score: 248.00 Matches: 92
 Percent Similarity: 35.07% Conservativeness: 49
 Best Local Similarity: 22.88% Mismatches: 143
 Query Match: 6.49% Indels: 118
 DB: 2 Gaps: 14

US-09-743-237-3 (1-2134) x F71410 (1-658)

QY 287 ATATATCCGGAAGAGCAACTTTGCAGAACTTCTTGTCTCAGAAATCTGTTGCAAGTTC 346
 Db 219 AsnGlyValGluLysGlnThrMetGlnHis----- 228
 QY 347 CCATGCTCCAGGAAGTACAGGATGCTCTGCTCTCTCTTTAAGAAAGATTCCAAACCA 406
 Db 229 -----AspSerAsnLysGluProGluSerAlaAsnAlaIleProTyrgluValAsnSer 246
 QY 407 ATGGTGATATGCCAA-----TTGAAGGGGGGCACACAAATGCTATGATAGAC 454
 Db 247 GlyValIleSerGlnAlaValSerLeuLeuHisArgGlyIleArgArgCysLeuAsp 266
 QY 455 -----AATTCTAGAACAAAGAACTA 475
 Db 267 PheGluMetProGlyAsnLysGlnThrSerSerGluAsnAsnThrAlaAlaCysGluSer 286
 QY 476 AAAGCACTCCATTTGGTTCCTCAGTAT----- 502
 Db 287 SerSerArgCysValProSerIleGlyLeuHisLeuAsnAlaIleLeuMetSerSer 306
 QY 503 CAAGATCAAAATAATTATCTACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGG 562
 Db 307 LysAspCysLysThrAsnValThrGlnAspTyrcysSerAlaAsnIleGlnValGly 326
 QY 563 AGATTTTGGCCAGCATCAACAAATTAATCTCATTACACAACTTGAAGGAGGCCCTTA 622
 Db 563 AGATTTTGGCCAGCATCAACAAATTAATCTCATTACACAACTTGAAGGAGGCCCTTA 622


```

Db 467 uLysArgTrpAsnAlaAspThrCysLysCysGluCys----- 479
QY 1348 AAATCTCTAGTTTAGAAAGGATGTTTAGGGGAACATGAGGCTGCTGCAGCAACACAC 1407
Db 480 -----GlnPr 481
QY 1408 AGGCTCCCTGCATCCCTGGCCAGGAGGAGTTTACTCA----- 1445
Db 481 oGlyMetProGlu---GlyCysGlyLys-GlnThrTrpIleSerAspLysCysLysC 500
QY 1446 --GAGCTCTCAGAGTGTGCACACCCATGCCCTTTCTGAGGAGGTGCATGCCCTGA 1503
Db 500 ysGluCysSerProThrIleThrCysGlnAlaProGlnIleLeuAspLeuAsnThrCysG 520
QY 1504 GCATTGTTTGT-----CTGGCCAGAGAGAGACTTGGGTTCCCATAGTCT 1551
Db 520 luCysLysCysProValAsnMetLeuAlaGlnLysGluLysCysLysSerProArgGlnT 540
QY 1552 GG-----GAGAGTGTCTGCAGGCGCGGAGGCGAGCAGCAGCGCCCTCGGAGAGCT 1602
Db 540 rpThrAspSerLysCysLeuCysGluCysSerThrThrProAlaThrCysGluGlyLysG 560
QY 1603 CACTCTGTCTCACTCTTCTCTCAGAGAAATGTTGCTCTGGAGGC 1646
Db 560 lnThrTrpCysGlyGluAlaCysGlnCysIleCysProGlyGly 574

RESULT 6
T00209
MEGF8 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00209
R:Nakajima, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T00209
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1737 <NAK>
A:Cross-references: EMBL:AB011541; NID:93449307; PIDN:BA032469.1; PID:93449308
A:Experimental source: brain; clone HGI392
C:Genetics:
A:Gene: MEGF8
A:Map position: 19q12

Alignment Scores:
Pred. No.: 0.0916 Length: 1737
Score: 121.00 Matches: 125
Percent Similarity: 29.04% Conservative: 60
Best Local Similarity: 19.62% Mismatches: 187
Query Match: 3.16% Indels: 265
DB: 2 Gaps: 35

US-09-743-237-3 (1-2134) x T00209 (1-1737)
QY 11 CAAGCGAAGCTCGCGGGGCGCAGCAGCGCGGAGCTCTCTCGGGAGTACCCCGG 70
Db 614 GlnGlyAlaLysArgAspArgMetArgAsnValArgGlySerSerArgGlyLeuGlyGln 633
QY 71 GATCCAGAGCTCAGCGCGCTGGAGAGCTCGCGCTCTCGAGGCGCCGCGAGCCGCCGCG 130
Db 634 ValPro-----GlyGluGlnProGlySerTrpGlyPheArgGluValArg 648
QY 131 CTGCACAGTGCACATT---CCTGTCTCTGCTGCTACCCGCGCAGCCGCGCGGGTGT 187
Db 649 LysLysMetAlaLeuTrpAlaAlaLeuAlaGlyThr-----GlyGlyPhe 663
QY 188 TTGCCCTCGGGCGCGCTGGTCTCT---CGAAGAGAGCTCTCCACCCGCGCGCTCCGATGA 244
Db 664 LeuGlu-----GluIleSerProHisLeuLysGluProArgProArgLeuPheHisAla 681
QY 245 TCCCATGTTGAATCAAGGTAAGCAGGTGGTACTACTACA-----AGTAATAAT 292

```

```

Db 682 SerAlaLeu-----LeuGlyAspThrMetValValLeuGlyGlyArgSerAsp 697
QY 293 CCGAAGAACGAACTTTTCAGAAATCTTCTGCTCAGGAATCCTGTTGCAAGTTCCCATGG 352
Db 698 ProAspGluPheSerSerAspValLeuLeuTyrGlnValAsnCys----- 712
QY 353 TCCAGGAACCTAGAGGATGCTCTCTCTTAAGAAAGATTCCAAACCAATGTGTG 412
Db 713 -----AsnAlaTrpLeu 716
QY 413 ATATGCCAAATTTGAAGGGGCACACAAATGTATGTATAGACAAATTTAGACAGAGAA 472
Db 717 LeuProAspLeu-----ThrArg--- 722
QY 473 CTAAGACACTCCATTTGGTTCTTCAGTATCAAGATCAAAATATATCTACAGTCACAT 532
Db 723 -----SerAlaSer 725
QY 533 GTCCTAAACCAACTGCTGCTTTAGTAGGAGATTTTTCAGCA-----TCAACAAATTA 589
Db 726 ValGlyProMetGluGluSerValAlaHisAlaValAlaValGlySerArgLeu 745
QY 590 AATCTCATACACAACAACTTGAAGGAGCCTTACCATCGGTAGTCAACGGTCTGCTTTC 649
Db 746 -----TyrIleSerGlyGlyPheGlyGlyValAlaLeuGlyArgLeuLeu 760
QY 650 CCCTCGGATCACTCTTCCAGGACCACCAAAATAACTTTGGCTGGTACTGTGACATGC 709
Db 761 -----AlaLeuThrLeuProProAspPro-----CysArgLeu 771
QY 710 TTTGCGAGTGGGACTTTTGCAACAAC-----TGCAAT---TGTAATAATTTGTTGC 757
Db 772 LeuSerSerProGluAlaCysAsnGlnSerGlyAlaCysThrTrpCysHisGlyAlaCys 791
QY 758 -----AACAACTTGCAT----- 769
Db 792 LeuSerGlyAspGlnAlaHisArgLeuGlyCysGlyGlySerProCysSerProMetPro 811
QY 770 -----CATGATATTGAACGGTTTAAAGCCATTAGGCTATGCTTTGGTAGAATCCA 820
Db 812 ArgSerProGluGluCysArgArgLeuArgThrCysSerGluCysLeuAlaArgHisPro 831
QY 821 GAAGCTTTCCAGCCCAAAATTTGGGAAGGCCAAATTTGGGCAATGTCAAGCCCGCAGCACA 880
Db 832 ArgThrLeuGlnPro-----GlyAspGlyGluAlaSerThr-----ProArgCysLys 847
QY 881 AAAGGTGCAACTCGCAGGAGGTGACGCTGCTG-----AAG 916
Db 848 TrpCysThrAsnCysProGluGlyAlaCysIleGlyArgAsnGlySerCysThrSerGlu 867
QY 917 AATTACTCGGAGTCTATGAGGCCAAATTTATGTTCTTCTTCTATTTC----- 964
Db 868 AsnAspCysArgIleAsnGlnArgGluValPheTrpAlaGlyAsnCysSerGluAlaAla 887
QY 965 -----AAATGCATTTGGTTCAGAAAT 985
Db 888 CysGlyAlaAlaAspCysGluGlnCysThrArgGluGlyLysCysMetTrpThrArgGln 907
QY 986 TATGAAGAAACCCAGAACGAAAGACATATGACATG-----CCAAACTAC----- 1033
Db 908 PheLysArgThrGlyGluThrArgArgIleLeuSerValGlnProThrTyrAspTrpThr 927
QY 1034 -----ATGCAGACTGGAGGTGTTGGAAGGAGCCATTTACCTG 1069
Db 928 CysPheSerHisSerLeuLeuAsnValSerProMetProValGluSerSerProProLeu 947
QY 1070 CCA-----CCAAGGAAATTTTCAGGACTTCCAGATTCAGTCAGATAGGCGCTTCCTCA 1126
Db 948 ProCysProThrProCysHisLeuLeuPro-----Asn 958
QY 1127 TGCATCTCTGGGAGGTGGTGGAGGCCACATGCGCTCGCTTGTCTCAGGAGAGAG 1186
Db 1127 TGCATCTCTGGGAGGTGGTGGAGGCCACATGCGCTCGCTTGTCTCAGGAGAGAG 1186

```

Db 959 CysThrSer-----CysLeuAspSerLysGlyAlaAsp 969
Qy 1187 GCCGAGAAAGAACACTGCTCCAGTGCCTGCCAGAGCAGATGATCTCGGAGGAATTTGGA 1246
Db 970 GlyGlyTrpGlnHisCys-----ValTrpSer 978
Qy 1247 AGTGCTTATCACAGATTCACACACTGAGTTTAAATCTAAGGATTTGAAATGGAGTAG 1306
Db 979 SerSerLeuGlnGlnCysLeuSerProSerTyr----- 989
Qy 1307 AGTATAAGTCTGAATGCTGATTTGCTTTAGTCTAGAAATCTCTAGTTAGAAAG 1366
Db 990 ---LeuProLeuArgCysMet----- 995
Qy 1367 GATGTTTAGGGGAACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCTGTCATCCCTG 1426
Db 996 -----AlaGly 997
Qy 1427 GGGCCAGGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTCTCTGA 1486
Db 998 GlyCysGlyArgLeuLeuArgGly----- 1005
Qy 1487 GGAGGTCATGGCTGAGCATTTGTTCTGCGCCAGAGAGAGAGCTTGGTTCCCATTA 1546
Db 1006 -----ProGluSerCysSerLeuGlyCysAlaGlnAlaThrGlnCysAlaLeu 1021
Qy 1547 GTCCTG-----GGAGAGTGTCTGCGAGGCGGC----- 1573
Db 1022 CysLeuArgArgProHisCysGlyTrpCysAlaTrpGlyGlnAspGlyGlyArg 1041
Qy 1574 -----GGAGGCGAGAGAGCGCTCGCGAG-----AGCTC 1603
Db 1042 CysMetGluGlyGlyLeuSer-GlyProArgAspGlyLeuThrCysGlyArgProGlyAl 1061
Qy 1604 ACTCTGCTGACTCTT-----CCTCTCAGAGAAATGTGCTCTCGA 1643
Db 1061 aserTrpAlaPheLeuSerCysProProGluAspGluCysAlaAsnGly 1077

RESULT 7
T00532
probable cadmium-transporting ATPase (EC 3.6.1.-) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999
C:Accession: T00532
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A:Reference number: 214167
A:Accession: T00532
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1001 <ROU>
A:Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176713
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 27/3; 113/1; 159/1; 268/2; 336/1; 403/3; 980/1
A:Note: T20K24.13
C:Superfamily: ATPase nucleotide-binding domain homology
C:Keywords: hydrolase
F:360-504/Domain: ATPase nucleotide-binding domain homology <ATN>

Alignment Scores:
Pred. No.: 0.12 Length: 1001
Score: 119.50 Matches: 110
Percent Similarity: 29.72% Conservative: 57
Best Local Similarity: 19.57% Mismatches: 184
Query Match: 3.12% Indels: 211
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x T00532 (1-1001)

Qy 235 GTCCGCATGATCCAGTTGAAATCAAGGTAAAGCAGGTAAAGCTGGTACTACTACAAGTAATATCC 294

Db 476 IleArgArgIleProGlnAlaValLysLeuAlaArg-----Arg 488
Qy 295 GGAGAAAGCAACTTTGTCAGAAATCTTCTGTCAGGAATCTGTTGCAAGTTCCCATGGTC 354
Db 489 AlaArgArgLysValValGluAsnValCysLeuSerIleIleLeu-Lys----- 504
Qy 355 CCAGGAACTAGAGATGCTCTGCTGCTCTCTCTTTAAGAAAGATTTCCAAACCAATGGT-- 412
Db 505 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProLeuIleTr 519
Qy 413 ---ATATGCCAATTGAAGGGGGGCACACAAATCTATGTATAGACAATTTCTAGA----- 463
Db 519 pAlaAlaValLeuValAspValGlyThrCysLeuValIleIlePheAsnSerMetLeuLe 539
Qy 464 -ACAAGAGAACTAAAGCACTCCATTTGGTTGCTTCAGTATCATCAAGATCAAAATAATTATCT 522
Db 539 uLeuArgGluLysLysLysIleGlyAsnLysLysCysTyrArg----- 553
Qy 523 ACAGTCAGATGTCCCTAAACCAATGACTGCTTTTAGTAGGAGAGATTTTGGCAGCATCAAC 582
Db 554 -----AlaSerThrSe 557
Qy 583 AAATTAATCTCATTTACACAACAACACTTGAGGAGCGCTTACCATCGGTAGTCAAC----- 637
Db 557 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyrValValAspLeuG 575
Qy 638 ---GGGTCTGCTTCCCTCGGACCAACTCTTCCAGGACCACCAAAAATAACTTTTGGC 693
Db 575 uAlaGlyLeuLeuThrLysSerGlyAsn----- 584
Qy 694 TGGTACTGTGACTGCTTTGCCAGTGGGAGCTTTTGCAACAACACTGCAAT----- 742
Db 585 -GlyGlnCysLysSerCysCysGlyAspLysLysAsnGlnGluValValMetMe 604
Qy 743 -----TGT----- 745
Db 604 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysCysGlyAspLysLy 624
Qy 746 -----AATAATTTGTTGC----- 757
Db 624 sGluGluLysValLysProLeuValLysAspGlyCysSerGluLysThrArgLysSe 644
Qy 758 -----AACAACTTGCATCATGATAT 777
Db 644 rGluGlyAspMetValSerLeuSerSerCysLysLysSerHisValLysHisAspLe 664
Qy 778 TGAACGCTTTAAAGCCATTAAAGCATCTCTTGGTAGAAATCCAGAACCTTTCCAGCCAAA 837
Db 664 u---LysMetLysGlySerGlyCysCysAlaSerLysAsnGluLysGlyLysGluVa 683
Qy 838 AATTGGGAAGGGCAATTTGGCAATGTCAAGCCCCCAGCACACAAGGGTGCAACTGCAG 897
Db 683 lValAlaLysSerCysCysGluLysProLysGlnValGlnValGluSerValGlyAspCysLy 703
Qy 898 GAGTCTAGGCTGCTGCTGAAG----- 916
Db 703 sSerGlyHisCysGluLysLysLysGlnAlaGluAspIleValValProValGlnIleI 723
Qy 917 ----AATTACTGGCAGTGTATGAGGCCCAATATATGTTCTTCTTCTTATTTGCAATGCAT 972
Db 723 eGlyHisAlaLeuThrHisValGluLeuGluGlnThrLysGluThrCysLysThrSe 743
Qy 973 TGGTTGCAAAAATTTATGAAGAAAGCCCCAGCAACGACATATGACATGCGCAAC-- 1030
Db 743 rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuSerSerGluAsnTh 763
Qy 1031 ----TACATCGACTGGAGGTTTG-----GAAGCAGCCCA 1062
Db 763 rProTyrLeuGluLysGlyValLeuIleLysAspGluGlyAsnCysLysSerGlySerGl 783
Qy 1063 TTACTCTCCCAACGAAATTTTCAGGACITTCCAAGATTCAGTCACCATAGCGGCGCTTC 1122

Db 783 uAsnMetGlyThrValLysGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh 803
QY 1123 CTCATGCATCTCCTGGGAGGTGGTGGAGGCCACATGCCCTGCCTGCTGCTCAGGAGA 1182
Db 803 rGlyGluLeuThr-----LeuAlaSerGluG1 812
QY 1183 AGAGCCGAGAAAGAACTACTCTCCAAAG-----TGCTGCGCAGAGAGATGCTCTGGA 1236
Db 812 uGluThrAspGlnAspCysSerSerGlyCysCysValAsnGluGlyThrValLysG1 832
QY 1237 GGAATTGGAGGTGCTTATCACAGATCTCCACACTGAGTTAAATCTAAGGATTCGA 1296
Db 832 nSerPheAspGluLysLysLysHisSerValLeu-----ValGluLysGluGlyLeuAs 849
QY 1297 AATGAGTAGAGTAAAGTGTGAATGCATGCTGATGTTGTTGCTTAGTCTAGAAATCTCTA 1356
Db 849 pMetGluThrGlyPheCysAspAlaLysLeuValCys----- 862
QY 1357 GTTTAGAAAGGATGTTTAGGGAACATAGAGCTGGCTCTGCAGCAACACAGGCTCCCC 1416
Db 862 ----- 862
QY 1417 TGCATCCTGGCCCGAGGAGTTTACTCAGAGCTCTCTGAAGATGTGCAACCCATGCC 1476
Db 863 ----- 866
QY 1477 CCTTTCTGAGGAGTGCATGGCCTGAGCAATGT---TGCTCTGGCCCGAGAGAGAGC 1533
Db 867 -----GluGlyGluValLysGluGlnCysArgLeuGluLeuLysLysGluGlu-- 882
QY 1534 TTGGTTCCCATAGTCTCGGAGAGTGTCTCAGCGCGCGGAGGCGCAGAGAGCCCTG 1593
Db 883 -----HisCysLysSerGlyCysGlyGluGluLeuGlnTh 895
QY 1594 CGGAGAGCTCACTGCTGCTGCTCTCTCAGAGATGT-----TGCTCTGAGGCT 1647
Db 895 rGlyGluLeuThrLeuValSerGluGluGlu-ThrGluSerThrAsnCysSerThrGlyC 915
QY 1648 GC 1649
Db 915 ys 915
RESULT 8
F84572
probable cadmium-transporting ATPase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84572
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: R84420; MUID:20083487; PMID:10617137
A:Accession: F84572
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1172 <STO>
A:Cross-references: GB:AE002093; NID:g4210504; PIDN:AAD12041.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g19110
A:Map position: 2
Alignment Scores:
Pred. No.: 0.121 Length: 1172
Score: 119.50 Matches: 110
Percent Similarity: 29.72% Conservative: 57
Best Local Similarity: 19.57% Mismatches: 184
Query Match: 3.12% Indels: 211
DB: 2 Gaps: 23
US-09-743-237-3 (1-2134) x F84572 (1-1172)

QY 235 GTCCGCATGATCCCATGTAATCAAGTAAGTGGTGGTACTACTACAAGTAATAATCC 294
Db 639 IleArgGileProGlnAlaValLysLeuAlaArg-----Arg 651
QY 295 GGAGAACAATTTGGCAGAACTCTTCTGCTCAGGAATCCTGTCAGAGTTCCCATGGTC 354
Db 652 AlaArgLysValValGluAsnValCysLeuSerIleIleLeu-Lys----- 667
QY 355 CCAGGAATAGAGGATGCTCCTGCTGCTTAAAGAAAGATTCACCAACCAATGGTG-- 412
Db 668 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProLeuIleTr 682
QY 413 ----ATATGCCAAATTGAAGGGGCACACAATGCTATATAGACAATTTCTAGA----- 463
Db 682 pAlaAlaValLeuValAspValGlyThrCysLeuLeuValIlePheAsnSerMetLeuLe 702
QY 464 -ACAAGAGAATAAAGACACTCCATTTGGTCTCAGTATCAAGATCAAAATAATATCT 522
Db 702 uLeuArgGluLysLysLysIleGlyAsnLysLysCysTyrArg----- 716
QY 523 ACAGTCAGATGTCCTAAACCAATGACTGCTTTAGTAGGAGATTTTCCCGAGCATCAAC 582
Db 717 -----AlaSerThrSe 720
QY 583 AAATAATAATCTCATACACAACAACTTGGAGGAGCTTACCATCGGTAGTCAAC----- 637
Db 720 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyrValValAspLeuG1 738
QY 638 ----GGGTCTGCTTCCCTCGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGC 693
Db 738 uAlaGlyLeuLeuThrLysSerGlyAsn----- 747
QY 694 TGGTACTGACTGCTTTGCCAGTGGGACTTTTGGCAACAACCTCAAT----- 742
Db 748 -GlyClnCysLysSerSerCysGlyAspLysLysAsnGlnGluAsnValValMetMe 767
QY 743 -----TGT----- 745
Db 767 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysCysGlyAspLysLy 787
QY 746 -----AATAATTTGTTGC----- 757
Db 787 sGluGluLysValLysProLeuValLysAspGlyCysSerGluLysThrArgLysSe 807
QY 758 -----AACAACTTGCATCATGATAT 777
Db 807 rGluGlyAspMetValSerLeuSerSerCysLysLysSerSerHisValLysHisAspLe 827
QY 778 TCAACGGTTTAAAGCCATTAAGGCATGTCTTGGTAGAAATCCAGAAGCTTTCCAGCCAAA 837
Db 827 u---LysMetLysGlyGlySerGlyCysCysAlaSerLysAsnGluLysGlyLysGluVa 846
QY 838 AATTGGGAAGGCCAAATTTGGGCAATGTCAAGCCCCAGCACACAAAGGTTGCAACTGCAG 897
Db 846 lValAlaLysSerCysCysGluLysProLysGlnGlnValGluSerValGlyAspCysLy 866
QY 898 GAGTCCAGCTGCTCTGAAG----- 916
Db 866 sSerGlyHisCysGluLysLysLysGlnAlaGluAspIleValValProValGlnIleI 886
QY 917 ----AATTACTCGAGTCTATGAGGCCCAATATGTTGTTCTTCTATTGCAATGAT 972
Db 886 eGlyHisAlaLeuThrHisValGluLeuGlnThrLysGluThrCysLysThrSe 906
QY 973 TGGTTGCAAAATTTATGAAGAAAGCCAGAACAGACACTAATGAGCATCCCAAC-- 1030
Db 906 rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuSerSerGluAsnTh 926
QY 1031 ----TACATGCAGACTGGAGGTTTG-----GAAGCGACCA 1062
Db 926 rProTyrLeuGluLysGlyValLeuIleLysAspGluGlyAsnCysLysSerGlySerG1 946
QY 1063 TTACCTGCCACCAACGAAATTTTTCAGGACTTCCAGGATTCACGATCAGTAGCGCCTTC 1122

QY 292 TCCGGAAGCAAC---TTTCAGAAATCTTCTGCTCAGGAATCTCTGTG---CAAGTT 345
 Db 702 GlyGlyArgAlaHisProGluThrProTyrSerGlySerGlyIleLeuLeuAspSerLeu 721
 QY 346 CCCATGGTCCCAAGAACTAGAGATGCCCTCCTGCTGCTCTTAAGAAAGATTCCACCC 405
 Db 722 ValLeuGlnProHisVal-----LeuMetLeu-----GluMetPheSerGly 735
 QY 406 AATGTTGATATGCCAATTGAAAGGGGCACACAAATGCTATATAGACAATCTAGAAC 465
 Db 736 GlyAspAlaAlaLeuGluArgThrThrPheGluArgTyrArg----- 751
 QY 466 AAGAGAACTAAAGCACTCCATTGCTTCTCCTCAGTATCAAGATCAAAATAATTACTACA 525
 Db 751 ----- 751
 QY 526 GTCAGATGTCCTAAACCAATCACTGCTTTAGTAGGAGATTTCGCCAGCATCAACAA 585
 Db 752 -----CysHisGluGluGlyLeu-MetProSerLysThrPr 763
 QY 586 ATTAATCTCATTAACACAACTTTCAGGGAGCCTTACCATCGGTAGTCAACGGGTCTGC 645
 Db 763 oLeuSer-----GluAlaCysValProLeuLeuIleSerAlaSerSe 777
 QY 646 TTTCCCTCGGGATCAACTTTCCAGGACCCACCAAAATAAATTGCTGGTACTGTGA 705
 Db 777 rLeuValTyrAsnGlyAlaLeuPro-----CysG1 787
 QY 706 CTGCTTTGCCAGTGGGACTTTTGCACAACTGCAATGTAATTAATTGTTGCAACAATT 765
 Db 787 nCysAspProGlnGlySerLeuSerSerGluCysAsnProHisGly----- 802
 QY 766 GCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAAATCCAGAAGC 825
 Db 803 -----GlyGlnCysArgCysLysProGlyValValGlyArgArgCysAspAl 818
 QY 826 TTTCCAGCAAAAATTGGGAAGGCCAATTGGGCAATGTCAAGCCGCCAGCAACAAAGG 885
 Db 818 a-----CysAlaThrGlyTyrTyrGly---PheGlyProAlaGlyCysGlnAl 833
 QY 886 GTGCAACTGCAGGAGTTCAGGCTGCTGAAGAAATTAATCTCGGATGCTATGAGGCCCAAT 945
 Db 833 aCysGlnCysSerProAspGlyAlaLeuSerAlaLeuCysGluGlyThrSerGlyGlnCys 853
 QY 946 TATGTTCTCTCT----- 958
 Db 853 sLeuCysArgThrGlyAlaPheGlyLeuArgCysAspHisCysGlnArgGlyGlnTrpG1 873
 QY 959 -----ATTGCAAA----- 967
 Db 873 yPheProAsnCysArgProCysValCysAsnGlyArgAlaAspGluCysAspAlaHisThr 893
 QY 968 -----TGCAATTTGTCAAAATTTATGAAGAAAGCCCAAGAAAGCAACACTAATGAG 1020
 Db 893 rGlyAlaCysLeuGlyCysArgAspPyr----- 902
 QY 1021 CATGCCAACTACATGCACACTGGAGGT-----TTGCAAGGCAGCCA 1062
 Db 903 -----ThrGlyGlyGluHisCysGluArgCysIleAlaGlyPheHi 916
 QY 1063 TTACCTGCCCAACCAAGAAATTTTCAGGACTTCCAGATTCAGTACGATAGGCGGCTTC 1122
 Db 916 sGlyAspProArgLeuProTyrGlyGlnCysArg----- 928
 QY 1123 CTCATGCATCTCTCGGAGGTGGTGGAGGCCACATCGCGCTGCTTGTCTCAGGAGGA 1182
 Db 929 -----ProCysProCys-----ProGluGlyPr 936
 QY 1183 AGAGGCCGAGAAAGAACTGCTCCAAAGTCCCTGGCAGCAGCAGATGATCTCGGAGGAAT 1242
 Db 936 oGlySerGlnArgHisPheAlaThrSerCysHisArgAsp----- 949
 QY 1243 TGGAAAGGTGCTTATCAGATTTCTCCACACTGAGTTTAAATCTAAGGATTGAAAATGGA 1302

Db 950 -GlyTyrSerGlnGlnIleValCysHisCysArgAlaGlyTyrThrGlyLeuArgCysG1 969
 QY 1303 GTAGAGTATAAAG----- 1317
 Db 969 uAlaCysAlaProGlyHisPheGlyAspProSerLysProGlyGlyArgCysGlnLeuCy 989
 QY 1318 TGAATGCCATGTTGATTTTCTCTAGTCTAGAAATCTCTAGTTAGAAAGGATGTTTAGGG 1377
 Db 989 sGluCysSer-----G1 993
 QY 1378 GAACATAGGCTGGCTCTGCAGAACACACAGGCTCCCTCGCATCCCTGGGCCCA----- 1432
 Db 993 yAsnIle-----AspProThrAspProGlyAlaCysAspProHisThr 1007
 QY 1433 -GGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTCTGAGAGG 1491
 Db 1007 rGlyGlnCysLeuArgCysLeuHisThrGluGlyPro----- 1020
 QY 1492 TGCATGGCTGAGCATTTGTTGCTGTGGCCAGAGGAGAGCTTGGGTTCCCATAGCTCCT 1551
 Db 1021 -----HisCys----- 1022
 QY 1552 GGGAGAGTCTGTCAGGCGCGGGAGGAGGAGC-----AGCCCTCGGAGAGCTCAC 1605
 Db 1023 -GlyHisCysLysProGlyPheHisGlyGlnAlaAlaArgGlnSerCysHisArgCysThr 1042
 QY 1606 T-----CTGGTCTGACTCTCTCTCAGAGAATGTTGCTCTGGAGGCTCTCTGCAATGAA 1659
 Db 1042 rCysAsnLeuLeuGlyThrAspProGlnArg-----CysProSe 1055
 QY 1660 AACCCCTAATGGTTTCTTGT 1678
 Db 1055 rThrAspLeuCysHisCys 1061
 RESULT 12
 T42215
 N:Alternate names: sperm-specific membrane protein
 C:Species: Mus musculus (house mouse)
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T42215
 R:GAO, Z.; Garbers, D.L.
 J. Biol. Chem. 273, 3415-3421, 1998
 A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane
 A:Reference number: Z22080; MUID:98123114; PMID:9452463
 A:Accession: T42215
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-5376 <GAO>
 A:Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1
 C:Genetics:
 A:Gene: Zan
 A:Map position: 5
 C:Function:
 A:Description: functions in multiple cell adhesion processes
 A:Note: found exclusively on the apical region of the sperm head
 C:Keywords: cell adhesion
 Alignment Scores:
 Pred. No.: 0.152 Length: 5376
 Score: 118.50 Matches: 137
 Percent Similarity: 29.39% Conservative: 64
 Best Local Similarity: 20.03% Mismatches: 232
 Query Match: 3.10% Indels: 251
 DB: 2 Gaps: 41
 US-09-743-237-3 (1-2134) x T42215 (1-5376)
 QY 16 CGAAGCTCCGGGGGGCGGAGCAGCGCGGGGAGCTCTCGGGGAGTACCCCGGATCC 75
 Db 3478 LysAspAlaGlnGlyAlaLeuLeuProAlaGlyLysThrTrpThrSerProGlyCysThr 3497


```

Alignment Scores:
Pred. No.: 0.164 Length: 2515
Score: 118.00 Matches: 132
Percent Similarity: 31.73% Conservative: 79
Best Local Similarity: 19.85% Mismatches: 240
Query Match: 3.09% Indels: 214
DB: 2 Gaps: 35

US-09-743-237-3 (1-2134) x S47008 (1-2515)

Qy 7 GGTCTAAGGCGAGCTCGGGGGGGCGACA--CGCACGGGGGGAGACTCTCTCGGGGACT 63
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 GlySerGlyGlySerAlaAlaThrValThrAlaThrSerAsnSerGlyThr 119
Qy 64 ACCCGGGATCCAGAGACTCAGCGCGCTGGAGAGCGTCGGCTCTGTCGAGGCCGCCGAC 123
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 AlaGlnGlyLeuGlnSer-----ThrSerAlaSerAlaGluAlaThrSer 134
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 124 CGCCCGCTCAACAGTGCACCTCTCTCTCGCTGCTACCCGCGCACCGCGCGGG 183
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 -----SerAlaAlaThrSer-----SerSerGlnSer 143
Qy 184 TGTTTTCGCCCTGGGGCGCTGGGTCTCGAAGAGCCTCCACCGGGCGTCGCGCATG 243
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144 SerLeuThrProSerLeuSerSerLeuAlaAsnAlaAsnGlyGlyAlaArgThr 163
Qy 244 ATCCAGTTCAATCAAGTAAAGTACGAGTGGTACTACTACAAGTAATAATCCGGAAGAC 303
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 PheProAla-ArgSerPheProAspGlyThrThrPheGly-----GlnIrl 179
Qy 304 AACTTTGCAGAAATCTTCTGTCTCAGCAATCCTGTTTCCAAGTTCCTCATGTCACGAGACT 363
||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 eThrLeuGlyGlnLeuThrLysGluIleGlnProTyrSerTyrTrpAsnMetGlnPh 199
Qy 364 AGAGGATGCCTCTGTCTTCTTTAAGAAGAT----- 397
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 199 eTyrGlnSerGluProAlaTyrValLysPheAspTyrThrIleProArgGlyAlaSerIrl 219
Qy 398 -----TCCAACCAATGTGTGATATGCCAATTCAAAGGGGCGCACAAAT 441
||| ||| ||||| ||||| ||||| ||||| ||||| |||||

```

Qy	122	GCGCCGCGCTGCACAGCTGCACACTTCTGTGTCCTCGCTGCTACTACC	CGGCCGACCG-----	172
Db	1296	SerSerAlaThrAlaProGluValProThrThrSerAlaSerThrProAspAlaVal	1315	
Qy	173	CAGCCGCGGGTGTT-----	-TTGCCCTGGGGCGCTGGGT	208
Db	1316	GluGluSerGlyIleProSerThrSerLysProThrAlaGluProLeuGluThrThrAla	1335	
Qy	209	CCTGCCGAAGGAG-----	---CCTCCC	226
Db	1336	ProSerThrGluValThrSerProGluGlySerGlyThrGluGluSerThrLeuProPro	1355	
Qy	227	ACCGGCGCTCCGATGATCCCATGTCGAATCAAGTAACGAGTGTACTACTACNA	286	
Db	1356	ThrGluGlySer-----	-GlyGluSerThrThrSer	1365
Qy	287	AATAATCCG-----	-GAAGAAGCAACTTTCAGAACTCTTCTGCTCAGGAATCCTGTTGC	340
Db	1366	SerAlaProThrValGluProAlaThrVal-----	-----	1375
Qy	341	AGTTCCCATGGTCCAGGAACACTAGAGGATGCTCTGCTGTTCTCTTAAGA	AAGATTTC	400
Db	1375	-----	-----	1375
Qy	401	AACCAANTGATATGCCAANTTGAAGGGGACACAAATGCTATGATAGACA	ATTCT	460
Db	1375	-----	-----	1375
Qy	461	AGAACAAGAACTAAAAGCACCTCCATTTCCTCAGTATCAAGATCAAAATA	TAT	520
Db	1376	-----	-LeuProGlnAsnArgAsnGluLys-----	1384
Qy	521	CTACAGTCAGATGTCCCTAAACCAATGACTGCTTTAGTAGGGAGATTTT	GCACAGCATCA	580
Db	1385	-----	-----ProGluPro	1387
Qy	581	ACAAATTAATCTCATTTACACAACTTCAGGAGCCCTACATCGGTAGTCA	ACGGG	640
Db	1388	ThrLysAspThrPhe-----	---AlaLeuprothrThrThrGly	1400
Qy	641	TCGTCTTCCCTCGGGATCAACTCTT-----	---	667
Db	1401	AlaProGlnAlaAsnAspSerSerValGluAsnThrLysCysThrSerSer	AspGluCys	1420
Qy	668	-----	---CCAGGA-----	673
Db	1421	GlyLeuAspAlaLeuCysGluArgArgThrGlyValcysArgCysGluProGly	PheGlu	1440
Qy	674	-----CCACAAAAATACTTTGGCTGGGTACTGTGACTGCTTTGGCACT	GGGACTTT	727
Db	1441	GlyAlaProProLysLysSerCysValaspValaspGluCys--AlaThrGly	Asp--	1458
Qy	728	TGCACAACTGCATTTGTAATAATTTGTCACAACTTGCATCATGATATTGA	ACGTTTT	787
Db	1459	--HisAsnCysHisGluSerAlaArgCysGlnAsn-----	---Tyr	1470
Qy	788	AAAGCCATTAAAGCATGCTTTGGTAGAAATCCAGAACCTTTCCAGCCAAA	AAATTTGGGAAG	847
Db	1471	ValGlyGlyThrAlaCysPhe-----CysProThrGlyPheArg--Lys	AlaAspAsp	1487
Qy	848	GGCCAATTGGCAATGTCAG-----CCCCAGCACAAAGAGGTGCAACTGC	AGGAGG	901
Db	1488	GlySerCysGlnaspIleAspGluCysThrGluHisAsnSerThrCys-----	-----	1503
Qy	902	TCAGGCTGCCTGAAGAATTACTGGAGTGTATGAGGCCCAAAATATTGTT	CTCTATT	961
Db	1504	-----	---CysGlyAlaAsn	1507
Qy	962	TGCAAAATGCATTGGT-----	-TGCAAAATTTGAAGAAAGC	997
Db	1508	AlalysCysValasnLysProGlyThrTySerCysGluCysGluAsn-----	-----	1523
Qy	998	CCAGAACAAGACACTAATGAGCATGCCAACTACATGCAGACTGGAGGTTT	GGAAGGC	1057

QY 321 CACAGTCAGATGTCTCCCTATACCCAAATGAGTCCTTTAGTAGGGAGAAITTTTGGCCAGCAICA 380
|||
P4 1205 P-GL-225 1207

1385	Db	-----ProGluPro	1386
581	Qy	ACAAATTAATCTCATTCACAACAACACTTGAGGGACCCTTACCATCGGTAGTCAACGGG	640
1388	Db	ThrLysAspThrPhe-----AlaLeuproThrThrThrGly	1400
641	Qy	TCGTCTTTCCCTCGGGATCAACTCTT-::: :::	667
1401	Db	AlaProGlnAlaAsnAspSerSerValGluAsnThrLysCysThrSerSerAspGluCys	1420
668	Qy	-----CCAGGA-	673
1421	Db	GlyLeuAspAlaLeuCysGluArgArgThrGlyValCysArgCysGluProGlyPheGlu	1440
674	Qy	-----CCACAAAATACTTTGGCTGGGTACTGTGCTGCCTTTGCCATGGGGGACTTT	727
1441	Db	GlyAlaProProLysLysSerCysValAspValaspluCys---AlaThrGlyAsp---	1458
728	Qy	TGCACCAACTGCATTTCTAATTGTTTGCACCAACTTGCATCATGATATTGAACGGTTTT	787

```
DB      I439  ---HISASHNCYSHISGUSERAIBARGCYSGINASH-----TYF  I470
```

Qy	788	AAAGCCATTAAAGCATGCTCTGGTAGAATTCCAGAACTTTCAGAGCCAAAAATTGGCAAG	847
Db	1471	ValGlyGlyTyrAlaCysPhe-----CysProThrGlyPheArg-----LysAlaAspAsp	1487
Qy	848	GGCCAATTGGCAATGTCAG-----CCCCAGCACAAACAGGGTGCAACTGCAGGAGG	901
Db	1488	GlySerCysGlnAspIleAspGluCysThrGluHisAsnSerThrCys-----	1503
Qy	902	TCAGGCTGCCTGAAGAAATTACTGCGAGTGCTATGAGGCCCAAAATTATGTTCTTCTTATT	961
Db	1504	-----CysGlyAlaAsn-----	1507
Qy	962	TGCAAAATGCATTGGT-----TCGAAAAATTATGAAGAAAGC	997
Db	1508	AlaLysCysValAsnLysProGlyThrTyrSerCysGluCysGluAsn-----	1523

```
Db 1524 -----||||||
Oy 1058 AGCATTAC-----CTGCCACCAACGAATTTTCAGGACTTCCAGATTCCAGTCACGAT 1111
Db 1528 AspGlyTyrGlnCysValProThrLys-----LysProCysAspSerThrGln 1544
Oy 1112 AGCGGCTTCCTCATGCTCTCTGGGAGTGTTGGAGGCCACATCGCTGCTGCTTT 1171
Db 1545 SerSerLysSerHisCys-----SerGluSerAsnMetSerCysGluVal 1559
Oy 1172 GCTCAGGAGAGAGCGCGGAGAACACTGCTCCAAAGTCCTGGCA----- 1219
Db 1560 AspThrValAspGlySerValGlu-----CysLysGluCysMetGlyGlyTyrLysLys 1577
Oy 1220 GAGCAGATGCTCGGAGGAAATTTGGAAGTGTCTTATCACAGATTCTCCACACTGAGTTT 1279
Db 1578 SerGlyLysValCysGluAspIleAsnGluCysValAlaGluLysAlaProCysSerLeu 1597
Oy 1280 AAATCTAAGGATTGAAAATGGAGTAGAGTATAAGTGTGAATGCATGTTGATTTGTCT 1339
Db 1598 AsnAlaAsnCysValAsnMetAsnGlyThrPheSerCysSerCys----- 1612
Oy 1340 TAGCTAGAAATCTCTAGTTAGAAAGAGATGTTAGGGGAACATGAGGCTGGCTCTCCAG 1399
Db 1613 -----LysGlnGly-TyrArgGlyAsp---GlyPheMetCysTh 1624
Oy 1400 C-----AACACACAGCTCCCTGCTCATCCCTGGGCCCGAGGAGTTTACTC 1444
Db 1624 rAspIleAsnGluCysAspGluArgHisProCysHisProHisAlaGluCysThrAsnLe 1644
Oy 1445 AGACTCTCTGAGATGTGGCAACCCATGCCCTTTTCT----- 1484
Db 1644 uGluGlySerPheLysCysGluCysHisSerGlyPheGluGlyAspGlyIleLysLysCy 1664
Oy 1485 -----GAGGAGTCCATGGCTG-----AG 1504
Db 1664 sThrAsnProLeuGluArgSerCysGluAspValGluLysPheCysGlyArgValAspHi 1684
Oy 1505 CATGTTTGTGTGCCCGCAGAGAGAGAGCTTGGTTCCCATAGTCTCGGAGAGTGCTG 1564
Db 1684 sValSerCysLeuSerValArgIleTyrAsnGly-----SerLeuSerSerValCy 1701
Oy 1565 CAGGCGCGGAGGCGAGAGCAGGCCCTGCGGAGAGCTACTGTGTCGACTCTCTCTCT 1624
Db 1701 sGluCysGluProGlyPheArgPheGluLysGluSerAsn-----SerCysValAspI 1719
Oy 1625 CAGAGATGT 1634
Db 1719 eAspGluCys 1722
```

Search completed: July 24, 2003, 14:36:58
Job time : 147.02 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:58:44 : Search time 78.0434 Seconds
(without alignments)
6494.675 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtcaggcggaag.....aaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 903798

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame-n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09743237/runat_24072003_113106_17768/app_query.fasta_1.4750
-DB=PublishedApplications_AA -QFAST=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USPR=US09743237@cgn_1.1.282@runat_24072003_113106_17768
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	730	19.1	147	11	US-09-764-891-3988	Sequence 3988, Ap
2	437	11.4	438	9	US-09-220-091-13	Sequence 13, Appl
3	278	7.3	53	15	US-10-106-698-6932	Sequence 6932, Ap
4	165	4.3	1057	15	US-10-189-971-6	Sequence 6, Appl
5	158	4.1	1192	15	US-10-189-971-18	Sequence 18, Appl
6	158	4.1	1207	15	US-10-189-971-20	Sequence 20, Appl
7	158	4.1	1477	15	US-10-189-971-8	Sequence 8, Appl
8	158	4.1	1535	15	US-10-189-971-14	Sequence 14, Appl
9	158	4.1	1593	15	US-10-189-971-4	Sequence 4, Appl
10	157.5	4.1	1251	15	US-10-189-971-16	Sequence 16, Appl
11	157.5	4.1	1342	15	US-10-189-971-24	Sequence 24, Appl
12	157.5	4.1	1512	15	US-10-189-971-10	Sequence 10, Appl
13	157.5	4.1	1570	15	US-10-189-971-12	Sequence 12, Appl
14	157.5	4.1	1628	15	US-10-189-971-2	Sequence 2, Appl
15	137	3.6	4123	15	US-10-213-509-5	Sequence 5, Appl
16	136	3.6	4588	15	US-10-137-129A-3	Sequence 3, Appl
17	130.5	3.4	2743	15	US-10-037-182-36	Sequence 36, Appl
18	130.5	3.4	3695	15	US-10-037-182-2	Sequence 2, Appl
19	125	3.3	1037	15	US-10-152-724A-3	Sequence 3, Appl
20	119	3.1	1036	11	US-09-887-527-60	Sequence 60, Appl
21	118.5	3.1	1801	10	US-09-938-275-8	Sequence 8, Appl
22	118	3.1	732	11	US-09-977-418-20	Sequence 20, Appl
23	118	3.1	732	11	US-09-977-033A-20	Sequence 20, Appl
24	117	3.1	2894	10	US-09-941-611-23	Sequence 23, Appl
25	117	3.1	2894	15	US-10-044-995-23	Sequence 23, Appl
26	117	3.1	3011	9	US-09-916-359-2	Sequence 2, Appl
27	117	3.1	3635	10	US-09-845-583-2	Sequence 2, Appl
28	117	3.1	3635	15	US-10-037-182-4	Sequence 4, Appl
29	116.5	3.0	1048	15	US-10-152-724A-4	Sequence 4, Appl
30	116	3.0	1036	15	US-09-373-967-4	Sequence 4, Appl
31	116	3.0	1036	15	US-10-028-072-142	Sequence 142, App
32	116	3.0	1036	15	US-10-121-049-142	Sequence 142, App
33	116	3.0	1036	15	US-10-123-904-142	Sequence 142, App
34	116	3.0	1036	15	US-10-140-470-142	Sequence 142, App
35	116	3.0	1036	15	US-10-175-746-142	Sequence 142, App
36	116	3.0	1036	15	US-10-176-918-142	Sequence 142, App
37	116	3.0	1036	15	US-10-137-865-142	Sequence 142, App
38	116	3.0	1036	15	US-10-140-474-142	Sequence 142, App
39	116	3.0	1036	15	US-10-142-431-142	Sequence 142, App
40	116	3.0	1036	15	US-10-143-114-142	Sequence 142, App
41	116	3.0	1036	15	US-10-140-002-142	Sequence 142, App
42	116	3.0	1036	15	US-10-142-419-142	Sequence 142, App
43	116	3.0	1036	15	US-10-123-262-142	Sequence 142, App
44	116	3.0	1036	15	US-10-142-423-142	Sequence 142, App
45	116	3.0	1036	15	US-10-142-423-142	Sequence 142, App

ALIGNMENTS

RESULT 1
US-09-764-891-3988
; Sequence 3988, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3988
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3988

Alignment Scores:

Pred. No.: 8,86e-54 Length: 147
Score: 730.00 Matches: 144
Percent Similarity: 97.30% Conservative: 0
Best Local Similarity: 97.30% Mismatches: 3
Query Match: 19,09% Indels: 1
DB: 11 Gaps: 0

US-09-743-237-3 (1-2134) x US-09-764-891-3988 (1-147)

QY 229 CCGGGCGTCCGATGATCCAGTCCAGTCAAGTAAAGCAGGTGGTACTACTACAGTAA 288
DB 1 ProGlyValArgMetIleProValGluLeuLysGlu-AlaGlyGlyThrThrSerAs 20
QY 289 TAATCCGGAAGAAGCACTTTCAGAACTTCTGCTCAGGAATCCTGTGCAAGTCCC 348
DB 20 nAsnProGluGluAlaThrLeuGlnAsnLeuAlaGlnGluSerCysLysPhePr 40
QY 349 ATGGTCCCAAGCACTAGAGTGCCTCTGCTCTCTTAAAGAAGATCCAAACCAAT 408
DB 40 oSerSerGlnGluLeuGluAspAlaSerCysSerLeuLysLysAspSerAsnProMe 60
QY 409 GGTGATATGCCAATGAAGGGGGGACACAAATGCTATATAGACAATTTCTAGAACAA 468
DB 60 tValIleCysGlnLeuLysGlyThrGlnMetLeuCysIleAspAsnSerArgThrAr 80
QY 469 AGAATCAAAAGCACTCCATTTGGTTCCTCAGTATCAAGATCAAAATAATTTATCTACAGTC 528
DB 80 gGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnTyrLeuGlnSe 100
QY 529 AGATGTCCTTAACCAAGCACTGCTTTAGTAGGGAGATTTTCCAGCATCAACAAT 588
DB 100 rAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLysLe 120
QY 589 AAATCTCATTACACAACAACTTGAGGAGCCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
DB 120 uAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAlaPh 140
QY 649 CCCCTCGGGATCAACTCTTCCA 670
DB 140 eProSerGlySerThr***Pro 147

RESULT 2

US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Geol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Alignment Scores:
Pred. No.: 1,15e-28 Length: 438
Score: 437.00 Matches: 92
Percent Similarity: 56.95% Conservative: 35
Best Local Similarity: 41.26% Mismatches: 62
Query Match: 11.43% Indels: 34
DB: 9 Gaps: 8

US-09-743-237-3 (1-2134) x US-09-220-091-13 (1-438)

QY 698 TACTGTGACTGCTTCCAGTCCAGTGGGACATTTTCAACAACACTCAATTTAATTTGTCG 757
DB 191 TyrCysAspCysPheAlaAsnGlyGluPheCysArgCysAsnCysLysAspCysHis 210
QY 758 AACAACTTGCATCATGATATTGAACGGTTTAAAGCATTAAAGCATCTCTTGGTAGAAT 817
DB 211 AsnAsnIleGluTyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 230
QY 818 CCAGAAGCTTCCACGCCAAAATTTGGG-----AAGGGCCAATTGGGCAATCTCAAGCCC 871
DB 231 ProAsnAlaPheLysProLysIleGlyIleAlaArgGlyGlyIleThrAspIleGluArg 250
QY 872 CAGCACAAACAAGGTCGCAACTGCAGAGGTGCTGCTGCTGAAGAATTTACTCGAGTGC 931
DB 251 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrCysGluCys 270
QY 932 TATGAGGCCAAATATGTTCTTCTTATTTCAATGCAATGCTTGGTTCGAAAAAT----- 985
DB 271 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 290
QY 986 -----TATGAAGAAAGC-----CCAGAACGAAAGACACTA 1015
DB 291 ThrTyrArgMetThrArgTyrLysAsnSerGlyAlaValSerAsnThrAsnAlaLeu 310
QY 1016 ATGAGCATGCCAAACTACATGACAGACT-----GGAGGTTTG----- 1051
DB 311 MetSerLeuThrAsnAlaSerSerThrAlaThrProAspSerGlyProGlySerValVal 330
QY 1052 -----GAAGCAGGCCATTACCTGCCACCAACAAATTTTTCAGGACTTCCAGATTC 1102
DB 331 ThrAspGluHisGlyAspAspTyr---GluAspMetLeuLeuSerHisLysProLysVal 349
QY 1103 AGTCACGATAGCGG-----CCTTCTCATCATCTCCTGGGAGTGTGGAGGCCACA 1156
DB 350 GluMetAspProArgPheProTyrTyrMetThrAspGluValGluAlaAla 369
QY 1157 TCGCGCTCGCTGCTTCCAGGAGGAGAGGCC-----GAGAAAGAA 1198
DB 370 ThrMetCysMetValAlaGlnAlaGluAlaLeuAsnTyrGluLysValGlnThrGlu 389
QY 1199 CACTGCTCAAGTGCCTGGCAGCAGCAGATGATCCTGGAGGAATTTGGAAGGTGCTTATCA 1258
DB 390 AspGluLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGlu 409
QY 1259 CAGATTCTC 1267
DB 410 GlnMetIle 412

RESULT 3

US-10-106-698-6932
; Sequence 6932, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6932

Alignment Scores:

Pred. No.: 2,43e-15 Length: 53
 Score: 278.00 Matches: 53
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.27% Indels: 0
 DB: 15 Gaps: 0

US-09-743-237-3 (1-2134) x US-10-106-698-6932 (1-53)

QY 1145 GTGGAGGCACATCGCTGCTGCTCTCAGGAGAGAGCGCGAGAAACACATGCC 1204
 Db ValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluAlaGluLysGlnHisCys 20
 QY 1205 TCCAGTCCCTGGCAGACAGATGATCTCTGGAGGAATTTGGAAGTGCTTATCAGATTT 1264
 Db SerLysCysLeuAlaGlnMetIleLeuGluPheGlyArgCysLeuSerGlnIle 40
 QY 1265 CTCACACTGAGTTTAAATCTAAGGATTTGAAATGGAG 1303
 Db LeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 53

RESULT 4

US-10-189-971-6
 ; Sequence 6, Application US/10189971
 ; Publication No. US20030028907A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Scoville, John
 ; APPLICANT: Turner, C. Alexander Jr.
 ; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-Like Proteins and Polynucleotides
 ; TITLE OF INVENTION: Same
 ; FILE REFERENCE: LEX-0360-USA
 ; CURRENT APPLICATION NUMBER: US/10/189,971
 ; CURRENT FILING DATE: 2002-07-03
 ; PRIOR APPLICATION NUMBER: US 60/302,949
 ; PRIOR FILING DATE: 2001-07-03
 ; PRIOR APPLICATION NUMBER: US 60/315,634
 ; PRIOR FILING DATE: 2001-08-29
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 1057
 ; TYPE: PRT
 ; ORGANISM: homo sapiens

US-10-189-971-6

Alignment Scores:
 Pred. No.: 2.3e-05 Length: 1057
 Score: 165.00 Matches: 171
 Percent Similarity: 27.67% Conservative: 62
 Best Local Similarity: 20.31% Mismatches: 226
 Query Match: 4.31% Indels: 383
 DB: 15 Gaps: 48

US-09-743-237-3 (1-2134) x US-10-189-971-6 (1-1057)

QY 16 CGAAGCTCGGGGGGCGACGACGAGCGGGGAGCTCTCGGGGAGTACCCCGGATCC 75
 Db ArgThrThrAlaAlaAlaValProLeuAlaGlyLys-----SerThrProAlaGlu 40
 QY 76 CAGAGCTCAGCGCGCTGGAGGACG----- 99
 Db ArgThrSerProThrProLeuThrProAlaValCysValAlaValSerGlyAsnValGln 60
 QY 100 -----TCGGCTCTCGAGCCCGCGCGCCCGC-----CCTGCAACG 138
 Db CysLeuAlaArgCysValProLeuProCysProGluProValLeuLeuProGlyGlu 80
 QY 139 TGCATTCTCTCGCTGCTTACCCGCGCAGCCCGCGGGGTGTTTGGCCCTGGG 198
 Db Cys-----CysProGlnCysProAlaAlaProAlaGlyCysProArgProGly 98

QY 199 GCG-----CCTGGGTCTCTGCGAAGGACCTCC 225
 Db AlaAlaHisAlaArgHisGlnGluTyrPheSerProProGlyAsp----- 113
 QY 226 CACCGCGCGTCCGATGATCCAGTGAAGTAAAGGTAAAGGTGGTACTACTACAAG 285
 Db ---ProCysArgArgCysLeuAspGlySerValSer----- 126
 QY 286 TAATAATCCGAAGAACCACTTTTCAGAACTTCTTCTCAGGAATCCTGTGTCAAGTT 345
 Db ---CysGlnArgLeuProCysProAlaProCysAlaHis---ProArgGlnGly 143
 QY 346 CCCATGGTCCAGGAAGTAGAGGAGCTCTCTG-----CTGTTCTCTTAAGNA 393
 Db ProCysCysProSerCysAspGlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGlu 163
 QY 394 AGATTCT---CAACCCCAATGGTATGCAATTTGAAAGGGGCGACACAATGCTATGAT 450
 Db ArgPheProSerProThr-AlaAlaCysHisLeu-----CysLeuCysTr 178
 QY 451 AGACAAATTTAGACAAAGAACTAAAGCAGCTC-----CATTTGGTTCTCT--- 496
 Db pGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCysProPheProAl 198
 QY 497 -----CAGTATCAAGATCAAAATAATTATCT 522
 Db aArgGlyAspCysCysProAspCysAspGlyCysGluTyrLeuGlyGlySerTyrLeu 218
 QY 523 ACAGTCAGATGTCCTTAAACCA-----ATGACTGCTTTAGTAGGAGATTT 567
 Db rasGlnGluPheProAspProArgGluProCysAsnLeuCysThrCysLeuGlyGlyPh 238
 QY 568 TTTGCCAGCATCAACAAATAATCTCATTTACACAACTTGAGGGAGCCTTACCATC 627
 Db eValThrCys-----GlyArgArgProCys 246
 QY 628 GGTAGTCAACGGTCTGCTTTCCCG----- 652
 Db sGluProProGlyCysSerHisProLeuLeuProSerGlyHisCysCysProThrCysGly 266
 QY 653 -----TCGGGATCAACTCTTCCAGGACCA----- 676
 Db nGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspProLeuAs 286
 QY 677 -CCAAAATAACTTTGGTGGTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 735
 Db pProThrCysSerLeu-----CysThrCysGlnGluGlySerMetArgCysGlnLy 303
 QY 736 C-----TCGAATTTGAATAA 750
 Db sLysProCysAlaProAlaLeuCysProHisProSerProGlyProCysPheCysProVa 323
 QY 751 TTGTTGCAACAACTTGCATCATGATPATTGAACGGTTTAAAGCCATTAAAGCATGCTTGG 810
 Db 1Cys-----HisSerCysLeuSe 329
 QY 811 TAGAATCCGAAGCTTTCCAGCCCAAAATTTGGGAAGGGCCCAATTTGGCAATGTCAGCC 870
 Db rGlnGlyArgGluHisGlnAspGlyGluGluPheGluGlyProAlaGlySerCysGlu-- 348
 QY 871 CCAGCACAAAGGCTGCAACTGCAGGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 930
 Db ---TrpCysArgCys 352
 QY 931 CTATGAGCCCAAAATTTATGCT----- 952
 Db sGlnAlaGlyGlnValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGly 372
 QY 953 -----TCTTCTATTGTC---AAATGCAATTTGGTTCACAAATTTAT---GAGA 993
 Db nValThrGluArgGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGly 392

```
QY 994 AAGCCAGACGAAGACACTAATGACATGCCAAACTACATGCAGACTGGAGTTTGGGA 1053
Db 1 |||
QY 392 uHisPro-----G 395
QY 1054 AGGAGCATTACCTGCCACCAACAAATTTTCAGGACTTCCAAGATTCAGTCACGATAG 1113
Db 1 |||
QY 395 uGlySerArgTrpValPro-----AspSe 404
QY 1114 GCGCCCTCTCTCATGCATCTCTCTGGAGGTGGTGGAGGCCACATCGGCC-----TG 1164
Db 1 |||
QY 404 rAlaCysSerSerCysValCysHisGluGlyVal---ValThrCysAlaArgIleGlnCy 423
QY 1165 CTG-----CTTGCTCAGGAGGAAGAGCGGAGAAACACACTGCTCCAGTGCCTGGC 1218
Db 1 |||
QY 423 sileSerSerCysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 441
QY 1219 AGACGATGATCTGGAGGAATTTGAAGGTGCTTATCACAGATTCTCCACATCAGTT 1278
Db 442 -----SerAspCy 444
QY 1279 TAAATCTAAGGATTGAAATGGAG-----TAGAGTATAAAGTG 1317
Db 444 sGluHisGluGlyArgIleGlyVal---ValThrCysAlaArgIleGlnCy 464
QY 1318 TGAATGATGTTGATTTTCTTCTAGCTAGTCTAGAAATCTCTAGTTTAGAAGGATTTAGGG 1377
Db 464 sGluValCysIle----- 468
QY 1378 GAACATGAGGTGCTCTGCAGCAACACAGGCTCCCTGCATCCCTGGCGCCAGGGAG 1437
Db 469 -----CysGluProGlnProGluGlyProSerLeu----- 479
QY 1438 TTTACTCAGAGCTCTCTGAAGATGT---GGCAACCCATGCCCCCTTTTCTGAGGA----- 1489
Db 480 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysPr 493
QY 1490 -----GGTGCATGCCCTGACCATGT----- 1510
Db 493 oProSerGlnLeuLeuProGlyProGlnHisCysCysProThrCysAlaGluAlaLe 513
QY 1511 -----TTGCTGGGCCA----- 1522
Db 513 userAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaProProaspProCysTrpTh 533
QY 1523 -----GAGGAGAGCTTGGTTCOCATA----- 1546
Db 533 rCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCy 553
QY 1547 -----GTCCCTGGGAGAGTGTCT 1563
Db 553 sProLeuSerGluArgHisThrProProGlySerCysCysProValCysArgGluCysVa 573
QY 1564 GCAGGGCGGGAGGAGGAGCAGGCGCCCTGGGAGAGCTCACTCTGCTGCACTTCTCTC 1623
Db 573 lVal-GluAlaGluGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProS 591
QY 1624 TCAGAGATGT-----TGCTCTGGAGGC-----T 1647
Db 591 erAsnAlaCysIleAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluGluC 611
QY 1648 GCTCTGATGAAACCCCTAATGTTCTTCTGTTTCTTCAATATTAGATAAGTT 1707
Db 611 ysGlnAla-----LeuSerCysProHisGlyTrpAlaLysValProG 625
QY 1708 CTCGGGATGGCTGTTGTGATACCCTTAAATCTCTAGAGAACTACTGAACACCTAAAG 1767
Db 625 lAlaAsp-SerCysCysGluArg----- 632
QY 1768 ATTTTCTGTACGCTAGATATTTCCTCCA-----GAGACACCGCAACTGTCTAGTCTTT 1818
Db 633 -----CysGlnAlaProThrGlnSerCysValHisGlnGlyArgGluVal----- 647
QY 1819 CCTAAGCCCCCGGAGACGACGCAATGGGGCTTCGAGGCCAGGCTTGACACCATG 1878
```

```
Db 648 -----AlaSerGlyGluArg-----TrpThrVal-----AspThrCysThrSerCys 661
QY 1879 TCTTGAGTTAGAGACTTAAATTTATCCAGTTCTTCTGTGTTTCTTACTTGAATGTGGA 1938
Db 662 SerCysMetalaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 681

RESULT 5
US-10-189-971-18
: Sequence 18, Application US/10189971
: Publication No. US20030028907A1
: GENERAL INFORMATION:
: APPLICANT: Walke, D. Wade
: APPLICANT: Turner, C. Alexander Jr.
: TITLE OF INVENTION: Same
: FILE REFERENCE: LEX-0360-USA
: CURRENT APPLICATION NUMBER: US/10/189,971
: PRIOR FILING DATE: 2002-07-03
: PRIOR APPLICATION NUMBER: US 60/302,949
: PRIOR FILING DATE: 2001-07-03
: PRIOR APPLICATION NUMBER: US 60/315,634
: PRIOR FILING DATE: 2001-08-29
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 18
: LENGTH: 1192
: TYPE: PRT
: ORGANISM: homo sapiens
US-10-189-971-18

Alignment Scores:
Pred. No.: 9.39e-05 Length: 1192
Score: 158.00 Matches: 169
Percent Similarity: 27.63% Conservative: 57
Best Local Similarity: 20.66% Mismatches: 219
Query Match: 4.13% Indels: 374
DB: Gaps: 46

US-09-743-237-3 (1-2134) x US-10-189-971-18 (1-1192)
QY 14 GCGAAGCTCGCGGGGGGCGACAGCGCGGGGAGTCTCTCGGGAGTACCCGGGAT 73
Db 198 GlycAlaPheGlyGlyLysGluTyProSerGlyAlaAspPheProHisProSerAsp 217
QY 74 CCAG-----AGCTCAGCGCGCTGGAGACGTCCGCTCTCGCAGG 114
Db 218 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 237
QY 115 CCGCGCAGCGCGCG-----CCTGCAACGTGCACCTCTCTGCTGCTGCTGTC 159
Db 238 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 255
QY 160 TACCGCGCAGCGCGCGCGGTGTTTCCCTCGGGCG----- 201
Db 256 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 275
QY 202 -----CCTGGTCTCTGCGAAGGACCTCCACCGCGCTCCGATCATCATC 246
Db 276 GluTyProPheSerProProGlyAsp-----ProCysArgArgCysLeu 289
QY 247 CCAGTTGAAATCAAGGTAAGCAGGTGGTGTACTACTACAGTAATAATCCGGAAGAACAC 306
Db 290 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 301
QY 307 TTGTCAGAACTCTCTTCTCAGGAATCCTGTCAGAACTCCCATGTTCCAGGACTAGA 366
Db 302 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 320
QY 367 GGATGCTCTCTG-----CTGCTCTTAAGAAAGATTC---CAACCAATGGT 411
Db 321 GlyCysLeuTyGlnGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 340
```

QY 412 GATATGCCAATTGAAGGGGGCACAAATGCTATGTATAGACAAATCTAGACAAAGACA 471
 Db 340 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysG1 355
 QY 472 ACTAAAGACATC-----CATTTGGTTCTCT----- 496
 Db 355 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 375
 QY 497 -----CAGTATCAAGATCAAAATATATCTACAGTACAGTGTCCCTAAACC 543
 Db 375 pCysaspGlyCysGluTrpLeuGlySerTyrLeuSerAsnGlnGluPheProAspPr 395
 QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTCCAGCAGCATCAACAAAATT 588
 Db 395 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 411
 QY 589 AAATCTCATACACAAACTTGAGGGAGCCCTTACATCGGTAGTCAACGGGTCTGCTTT 648
 Db 412 -----GlyArgArgProCysGluProProGlyCysSerH1 423
 QY 649 CCCC----- 652
 Db 423 sProLeuLeuProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 443
 QY 653 -----TCGGGTCAACTCTTCCAGGAGCCACCAAAATAAATCTTGGCTGGGTACTG 702
 Db 443 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 458
 QY 703 TGACTGCTTTGCCAGTGGGAGCTTTTGGCAACACTGCAATGTATAATTTGTCACAA 762
 Db 459 -----CysSerLeuCysThrCysGln-----GlyAr 467
 QY 763 CTTGCATCATGATTTGAACGTTTAAAGCCATTAAAGCATCTCTTGGTAGAAATCCAGA 822
 Db 467 gGluHisGlnAspGlyGluGluPhe----- 475
 QY 823 AGCTTTCCAGCCAAAATTTGGGAAGGGCCAATTTGGGCAATGTCAAGCCCCAGCACAA 882
 Db 476 -----GluGlyProAlaGlySerCysGlu----- 483
 QY 883 AGGTGCAACTGCAGAGGTCTAGGCTGCCTGAAGAAATTAAGTGTATGAGGCCCA 942
 Db 484 -----TrpCysArgCysGlnAlaGlyG1 491
 QY 943 AATTATGTCT----- 952
 Db 491 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 511
 QY 953 -TCCTTCTATTGC---AAATGCATTGGTTGCCAAAATTTAT---GAAGAAAGCCCAAGG 1005
 Db 511 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 529
 QY 1006 AAAGACACTAATGAGCATGCCAACTACATGACAGCTGGAGGTGGGAAGGAGCCATTA 1065
 Db 530 -----GluGlySerArgTr 534
 QY 1066 CTGCGCACCACAAATTTTCAGGACTTCCAGATTCCAGATTCAGTACGATAGCGGCTCTCTC 1125
 Db 534 pValProPro-----AspSerAlaCysSerSe 543
 QY 1126 ATGCATCTCTGGGAGTGGTGAGGCCACATGGCC-----TGCTGT-----CT 1170
 Db 543 rCysValCysHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 562
 QY 1171 TGCTCAGGAGAGAGCGCCAGAAAGAACACTGCTCCAAAGTCCCTGGCAGCAGATGAT 1230
 Db 562 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 576
 QY 1231 CTGGAGGAATTTGGAGGTGCTTATCACAGATTCTCCACACTGAGTTAAATCTAAGG 1290
 Db 577 -----SerAspCysGluHisGluG1 583

QY 1291 ATTGAAATGGAG-----TAGACTATAAAGTGTGAATGCATGTT 1329
 Db 583 yArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysI1 603
 QY 1330 GATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAAGATGTTTAGGGGAACATGAGGCT 1389
 Db 603 e----- 603
 QY 1390 GGCTCTGCAGCAACAACAGAGTCCCTGCATCCCTGGGCCCGAGGAGTTTACTCAGAGC 1449
 Db 604 -----CysGluProGlnProGluGlyProProSerLeu----- 614
 QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCCCTTTTCTCAGGA----- 1489
 Db 615 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 632
 QY 1490 -----GGTGCATGGCTGAGCATGT----- 1510
 Db 632 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 652
 QY 1511 -----TTGTCTGGCCCA-----CA 1524
 Db 652 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysG1 672
 QY 1525 GGAGAGAGCTTGGTTCCTCCATA----- 1546
 Db 672 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerG1 692
 QY 1547 -----GTCTGGGAGAGTGTCTGCAGGGGGGGGG 1575
 Db 692 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaG 712
 QY 1576 AGGCGAGAGCGCCCTCGGGAGAGCTCACTGTGTCGACTCTTCTCTCAGAGATGT- 1634
 Db 712 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 730
 QY 1635 -----TGCTCTGGAGGC-----TGCTCTGCATGAA 1659
 Db 730 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla--- 748
 QY 1660 AACCTAATGGTTCTTGTGTTTCAATATTTAGAAATAGTTCTCCGGATGGGC 1719
 Db 749 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 763
 QY 1720 TGTGTGTATACCATTTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTCTCTAGC 1779
 Db 764 CysCysGluArg-----CysGln 769
 QY 1780 GTAGATATTTCGCCA-----GAGACACGCAACTGTCACTCTTCTTAAGGCCCC 1830
 Db 770 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 784
 QY 1831 GGGAGACGAGCAATGGGCGCTCGCAGGCCAGGTTCACCAGCATGTCTTGAGTTAGA 1890
 Db 785 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMeta1a 800
 QY 1891 GCACITAAATATTCAGTTTCTTCTGTCTTCTTCTTCTCAATGTGGA 1938
 Db 801 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 816

RESULT 6

US-10-189-971-20
 ; Sequence 20, Application US/10189971
 ; Publication No. US20030028907A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walker, D. Wade
 ; APPLICANT: Scoville, John
 ; APPLICANT: Turner, C. Alexander Jr.
 ; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucle
 ; TITLE OF INVENTION: Same
 ; FILE REFERENCE: LEX-0360-USA
 ; CURRENT APPLICATION NUMBER: US/10/189,971
 ; CURRENT FILING DATE: 2002-07-03

; PRIOR APPLICATION NUMBER: US 60/302,949
 ; PRIOR FILING DATE: 2001-07-03
 ; PRIOR APPLICATION NUMBER: US 60/315,634
 ; PRIOR FILING DATE: 2001-08-29
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 20
 ; LENGTH: 1207
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-189-971-20

Alignment Scores:

Pred.. No.: 9,42e-05 Length: 1207
 Score: 158.00 Matches: 169
 Percent Similarity: 27.63% Conservative: 57
 Best Local Similarity: 20.66% Mismatches: 219
 Query Match: 4.13% Indels: 374
 DB: 15 Gaps: 46

US-09-743-237-3 (1-2134) x US-10-189-971-20 (1-1207)

```

QY 14 GCGNAGCTCGGGGGCGACGGAGCGGGAGCTCTCGGGGAGTACCCGGGAT 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 213 GlyCysAlaPheGlyGlyLysGluTyrProSerGlyAlaAspPheProHisProSerAsp 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 74 CCCAG-----AGCTCAGCGCGCTGGAGGACGTCGCGCTCTCGCAGG 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgArgCysVal 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 115 CCCGCGACGCCCG-----CCTGCACGTCGACTTCTGTCCTCGCTGC 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 TACCGCGCCAGCGCGCGGGTGTTCCTCCCTCGGGCG----- 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 -----CCTGGTCTCGGAAGAGCGCTCCACCGGGCGTCCCGATGATC 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 GluTyrPheSerProGlyAsp-----ProCysArgArgCysLeu 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 CCAGTTGAATCAGGTAAGCAGGTGGTACTACTACAGTAATTAATCCGGAAGAACAC 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 TTTCAGAAATCTCTTGCTCAGGAATCTGTTGCAAGTTCCATGGTCCCGAGGAACAGA 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 CysProProAlaProCysAlaHis----ProArgGlnGlyProCysCysProSerCysAsp 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 GGATGCCCTCTG-----CTGPTCTCTTAAGAAAGATTC---CAACCCCAATGGT 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 412 GATATGCCAANTGAAGGGGCGACACAATGCTATGTATAGACAAATCTTAGAACAGAGA 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysG1 370
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 472 ACTAAAGCACTC-----CATTTGGTTCCP----- 496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 390
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 497 -----CAGTATCAAGATCAAAATAATATCTACAGTCAGATGTCCCTAAACC 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 pCysAspGlyCysGluTyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspPr 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 544 A-----ATGACTGCTTTAGTAGGAGGATTTTGGCAGCATCAACAAAT 588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 AAATCTCATACACAACAACTTGAGGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 -----GlyArgProCysGluProProGlyCysSerHi 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

QY 649 CCCC----- 652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 653 -----TCGGGATCAACTCTTCAGGACCAACAAAAATAACTTTGGCTGGGTACTG 702
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 703 TGACTGCTTCCCACTGGGACCTTTTGCACAACTGCAATTTGTAATTAATTTTGCACAA 762
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 -----CysSerLeuCysThrCysGln-----GlyAr 482
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 763 CTTCATCATGATATTGAACGGTTTAAAGCCATTAAAGGCATGCTCTGTAGAAAAATCCAGA 822
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 gGluHisGlnAspGlyGluGluPhe----- 490
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 823 AGCTTCCAGCCAAAAATTTGGGAAGGCGCAANTTGGCAANTTCAAGCCCCAGCACAA 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 -----GluGlyProAlaGlySerCysGlu----- 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 883 AGGCTGCAACTGCAGGAGTCCAGCTGCCTCAAGAAATTTACTCGAGTCTATGAGGCCA 942
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 499 -----TrpCysArgCysGlnAlaGlyG1 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 943 AATTATGTGT----- 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 953 -TCCTTATTTCG---AAATGCATTGGTTGCAAAATAT---GAAGAAACCCAGAACG 1005
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1006 AAAGACACTAATGAGCATGCCAACTACATCCAGACTGGAGGTTTGGGAAGCAGCCATTA 1065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 -----GluGlySerArgTr 549
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1066 CCTGCCACCAACAAATTTTTCAGGACTTCCAAAGATTTCAGTCACGATAGCGGCTTCCTC 1125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 pValProPro-----AspSerAlaCysSerSe 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1126 ATGCATCTCTGGGAGGTGGAGGCCACATGCGCC-----TCGCTG-----CT 1170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 rCysValCysHisGlyGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1171 TGCTCAGGAGAGAGAGCCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGAT 1230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1231 CCTGGAGGAATTTGGAAGGTGCTTATCAGAGTTTCCACACTCTCCACACTAGTTAAATCTAAGG 1290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 -----SerAspCysGluHisGluG1 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1291 ATTTAAATCGAG-----TAGAGTATAAGTGTGAATGCATGTT 1329
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 yArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIl 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1330 GATTTTGTCTTAGTCTAGAAATCTCTAGTTTGAAGAGGATCTTTAGGGGAACATGAGGCT 1389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 e----- 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1390 GGCTCTCAGCAACAAACAGGCTCCCTGTGCATTCCTGGGCCAGGGAGTTTACTCAGAGC 1449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 ---CysGluProGlnProGluGlyProProSerLeu----- 629
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1450 TCTCTGAGATGT---GGCAACCCATGCCCTTTTCTGAGGA----- 1489
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 630 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 647
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1490 ---GGTGCATGGCTGAGCATTTGT----- 1510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 uLeuProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 667
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

QY 1511 -----TTCTCTGGCCCA-----GA 1524
Db 667 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProcysTyrThrCysGlnCysG1 687
QY 1525 GGAGAGAGCTGGGTCCCAT-
Db 687 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerG1 707
QY 1547 -----GTCTCTGGGAGAGTCTCTCGAGGGGGGG 1575
Db 707 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaG 727
QY 1576 AGGGCAGAGAGCCCTCGGAGAGACTACTCTGTCTGCTGCTCTCTCTCAGAGAAATGT- 1634
Db 727 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 745
QY 1635 -----TGCTCTGGAGGC-----TGCTCTGCATGNA 1659
Db 745 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluCysGlnAla---- 763
QY 1660 AACCCCTAATGTTCTGTTTGTCTTCAAAATATTAGAAATAAGTTCTCTCGGATGGGC 1719
Db 764 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 778
QY 1720 TGTGTGTATACACTTAAATCTCTAGAGAACTACTGACACCTAAAGATTTTCTGTAGC 1779
Db 779 CysCysGluArg-----CysGln 784
QY 1780 GTAGATATTCCCCA-----GAGACACCGCAACTGTCAGTCTTTCCTTAAGCCCCC 1830
Db 785 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 799
QY 1831 GGGAGACGAGGCAATGGGCGCTCGCAGCCAGGCTTGCCAGCAGCATGCTTTGAGCTAGA 1890
Db 800 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMetala 815
QY 1891 GGACTTAAATATCATCTCTCTGTTCTTCTACTTGAATGTGGA 1938
Db 816 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 831

RESULT 7

US-10-189-971-8
; Sequence 8, Application US/10189971
; Publication NO. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kieelin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-8

Alignment Scores:

Pred. No.: 9,93e-05 Length: 1477
Score: 158.00 Matches: 169
Percent Similarity: 27.63% Conservative: 57
Best Local Similarity: 20.66% Mismatches: 219
Query Match: 4.13% Indels: 374
DB: 15 Gaps: 46

US-09-743-237-3 (1-2134) x US-10-189-971-8 (1-1477)
QY 14 GCGAAGCTCGGGGGGGGCGGACGACGCGGGGAGCTCCTCGGGAGTAGTACCCCGGAT 73
Db 483 GlyCysAlaPheGlyGlySerGluTyrProSerGlyAlaAspPheProHisProSerAsp 502
QY 74 CCCAG-----AGCTCAGCGCTGGAGGACGCTCGCGCTCTCTCTCTCTCTCTCT 114
Db 503 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgArgCysVal 522
QY 115 CCCCGCAGCCCGCCG-----CCTGCAAGCTGCACCTCTCTCTCTCTCTCTCTCT 159
Db 523 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 540
QY 160 TACCCGCGCAGCGACCGCGGGGTGTTTGGCCCTGGGGCG----- 201
Db 541 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 560
QY 202 -----CCTGGGTCTCGGAAGAGGCTCCCAACCGGGGCTCGCATGATC 246
Db 561 GluTyrPheSerProProGlyAsp-----ProCysArgArgCysLeu 574
QY 247 CCAGTTGAANTCAAGTAAGCAGGTGCTACTACTACAAGTAAATATCCGGAAGAGCAAC 306
Db 575 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 586
QY 307 TTTGCGAAGATCTTCTGCTCAGGAATCTGTTGCAAGTTCCTCATGTCCTCGGAGACTAGA 366
Db 587 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 605
QY 367 GGATCCCTCCTG-----CTGTTCTCTTAAGAAAGATTC---CAACCCCAATGGT 411
Db 606 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 625
QY 412 CATATGCCAATTTGAAGGGGCGCACAAATGCTATGTATAGCAATTTCTAGAACAGAGA 471
Db 625 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysG1 640
QY 472 ACTAAAGACACTC-----CATTTGTTCTCT----- 496
Db 640 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyArgCysCysProAs 660
QY 497 -----CAGTATCAGATCAAAATAATTATCTACAGTCAGATGTCCTTAAACC 543
Db 660 pCysAspGlyCysGluTyrLeuGlyLysSerTyrLeuSerAsnGlnGluPheProAspPr 680
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTTCAGCAGATCAACAAAAT 588
Db 680 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 696
QY 589 AAATCTCATACACAACACTTTGAGGAGCCCTTACCATCGGTAGTCAACGGGCTCTCTTT 648
Db 697 -----GlyArgArgProCysGluProProGlyCysSerH1 708
QY 649 CCCC----- 652
Db 708 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 728
QY 653 -----TCGGGATCAACTCTTCCAGGAGCCACCACCAATAAATACTTTGGCTGGGTACTG 702
Db 728 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 743
QY 703 TGACTGCTTTGCCAGTGGGAGCTTTTGAACAACCTGCAATTCATAATAATTGTTGCAACAA 762
Db 744 -----CysSerLeuCysThrCysGln-----GlyAr 752
QY 763 CTTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATCTCTGTTAGAAATCCAGA 822
Db 752 gGluHisGlnAspGlyGluGluPhe----- 760
QY 823 AGCTTTCCAGCCAAAANTTGGGAAGGGCCAATTTGGCAATGTCAAGCCCAACACACAA 882
Db 761 -----GluGlyProAlaGlySerCysGlu----- 768

756 "proI vsAlaC vsAla proAlaI enC vsD proPh e proAlaA r nG l vA s nC vsC vs proAs 776

497	-----CAGTATCAAGATCAAAATAATTTATCTACAGTCAGATGTCCTAAACC	543
	::: ::	
776	pCysAspGlyCysGluThrLeuGlyCysSerIleuSerAsnGlnGluPheProAspPr	796
	::: ::	::: ::
544	A-----ATGACTGCTTTAGTAGGAGATTTTTCGACGAGCATCAACAAATTT	588
	::: ::	: ::
796	oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys-----	812
589	AAATCTCATTTACACAACAATTGAGGGAGCCCTTACCATCGGTAGTCAACGGGTCTGCTTTT	648
813	-----GlyArgArgProCysGluProProGlyCysSerHi	824
649	CCCC	-----

824	sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgIleHisGlyVa	844
653	-----TCGGGATCAACTCTTCCAGGAGCCACCACCAATAACTTTGGCTGGGTACTG	702
844	lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr-----	859
703	TGACTGCTTTGCCAGTGGGGACTTTTTCGCAACAACTTTCGTAATAATTCGTGCACAA	762
	: ::	

800		CysSerLeucysIleCysGlu	GlyAla	800
763	CTTGCATCATGATATTGAACGGTTTAAGCCCAATTAAAGCATGTCTTGTTAGAATAATCCAGA			822
868	gGluHisGlnAspGlyGluGluPhe			876
823	AGCTTTCCAGCCAAAAAATTGGGAAGGGCCAATTGGGCAATGTCAAGCCCCAGCACAA			882
877	-----GlUGlyproAlaGlySerCysGlu-----	:::::		884

[illegible]


```

QY 1390 GCCTCTGAGCAACACAGGCTCCCTGCATCCCTGGCCAGGAGTTTACTCAGAGC 1449
Db 1005 ---CysGluProGlnProGluGlyProProSerLeu----- 1015
QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCTTTCTGAGGA----- 1489
Db 1016 ---ArgCysHisArgGlnCysProSerLeuValGlyCysProProSerGlnLe 1033
QY 1490 ---GGTCATGGCTGAGCATTGT----- 1510
Db 1033 uLeuProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 1053
QY 1511 ---TTGCTGGGCCA-----GA 1524
Db 1053 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGl 1073
QY 1525 GGAGAGACTGGGTTCCTCATA----- 1546
Db 1073 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGl 1093
QY 1547 ---GTCTCTGGAGAGTCTCTGCAGGGCGCGG 1575
Db 1093 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaG 1113
QY 1576 AGGCAGAGCAGGCCCTCGGAGAGCTCACTTGTCTGACTCTTCTCTCAGAGAATGT- 1634
Db 1113 LuGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 1131
QY 1635 ---TGCTCTGGAGC----- 1659
Db 1131 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluCysGlnAla---- 1149
QY 1660 AACCCTAATGTTCTTGTGTTTTCATAATTATTTAGAAATAAGTTCTCCGGATGGC 1719
Db 1150 ---LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 1164
QY 1720 TGTGTGTATACACTTAAATCTCTAGAGAACTACTGACACCTAACACATTTTCTGTAGC 1779
Db 1165 CysCysGluArg-----CysGln 1170
QY 1780 GTAGATATTCCCCA-----GAGACACCGCACTGTCTAGTCTTTCTTAAGCCCCC 1830
Db 1171 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 1185
QY 1831 GGGAGACGACGACCAATGGGCTCGCAGGCCAGGCTTGACAGCATGTCTTGACTTAGA 1890
Db 1186 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMetAla 1201
QY 1891 GGACTTAAATATACAGTTTCTCTGTGTTTCTACTTGAATGTGGA 1938
Db 1202 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1217

```

RESULT 10

```

US-10-189-971-16
; Sequence 16, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16

```

```

; LENGTH: 1251
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-16

Alignment Scores:
Pred. No.: 0.000105 Length: 1251
Score: 157.50 Matches: 172
Percent Similarity: 27.78% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 231
Query Match: 4.12% Indels: 373
Db: 15 Gaps: 47

```

```

US-09-743-237-3 (1-2134) x US-10-189-971-16 (1-1251)
QY 14 GCGAAGCTCGCGGGGGGCGACGACGCGGGGGGCTCTCGGGAGTACCCCGGAT 73
Db 222 GlyCysAlaPheGlyLysGluTyrProSerGlyAlaAspPheProHisProSerAsp 241
QY 74 CCCAG-----AGCTCAGCGCTCGAGGACGCTCGCGCTCTCCAGG 114
Db 242 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 261
QY 115 CCCCCAGCGCGCCG-----CCTGCAAGCTGCACCTTCTCTGCTCGCTGCG 159
Db 262 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 279
QY 160 TACCCGCGCACCAGCGCGCGGTGTTTGGCCCTGGGGCG----- 201
Db 280 ProAlaAlaProAlaAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 299
QY 202 -----CCTGGTCTCTCGAAGGAGCTCCACCCGCGGCTCGCATGATC 246
Db 300 GluTyrPheSerProGlyAsp-----ProCysArgArgCysLeu 313
QY 247 CCAGTTGAATCAAGTAAGCAGGTGTACTACTACAAGTAATAATCCGGAAGAGCAAC 306
Db 314 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 325
QY 307 TTTGCAGAATCTTCTGTCAGGAATCTGTTGCAAGTTCCCATGTGCCAGGAAGTAGA 366
Db 326 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysasp 344
QY 367 GGATCCCTCTG-----CTGTTCTCTTAAGAAAGATTC---CAACCCCAATGTT 411
Db 345 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 364
QY 412 GATATGCCAATTCGAAGGGGCGACAAATGCTATGTATAGACAATTTAGAACAGAGA 471
Db 364 alaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysgl 379
QY 472 ACTAAAGCAGCTC-----CATTTGTTCTCT----- 496
Db 379 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProms 399
QY 497 -----CAGTATCAAGATCAAAATAATTTACTACAGTCAGATGTCCTCAACCC 543
Db 399 pCysaspGlyCysGluTyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspPr 419
QY 544 A-----ATGACTGCTTTAGTAGGGAGATTTTTCAGCAGCATCAACAAAT 588
Db 419 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 435
QY 589 AAATCTCATACAAACAACTTTCAGGAGCCTTACCATCGGTAGTCAACGGGTCTCTTT 648
Db 436 -----GlyArgArgProCysGluProProGlyCysSerHis 447
QY 649 CCCC----- 652
Db 447 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 467
QY 653 -----TCGGGATCAACTCTTCCAGGACCA-----CCAAAAATAACTTTGCTGG 696

```

Db 467 lThrThralaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----- 485
Qy 697 GTACTGTGACTGCTTTGCCAGTGGGACTTTTGCACAAAC----- 736
Db 486 ----CysThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLe 504
Qy 737 -----TGAATTGTAAATAATTGTTCACAACTTGCATCA 771
Db 504 uCysProHisProSerProGlyProCysPheCysProValCys----- 518
Qy 772 TGATATTGAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAATCCAGAACTTTCCA 831
Db 519 -----HisSerCysLeuSerGlnGlyArgGluHisGlnAs 530
Qy 832 GCCAAAATTTGGGAGGGCCAAATGGCAATGTCAAGCCCGCAGCACAAAGGGTGCAA 891
Db 530 pGlyGluLupheGluGlyProAlaGlySerCysGlu----- 542
Qy 892 CTGCAGGAGTCAAGCTGCTGGAAGAAATTACTCGAGTGTATGAGGCCCAAAATTATGTG 951
Db 543 -----TrpCysArgCysGlnAlaGlyGlnValSerCy 553
Qy 952 T-----TCCTCTAT 960
Db 553 svalArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCy 573
Qy 961 TTGC---AAATGCATTGTGTCACAAAATTAT---GAAGAAAGCCAGAACGAAAGACACT 1014
Db 573 sCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 588
Qy 1015 AATGAGATGCCAAACTACATGCAGATGGAGGTTTGGAGGCGCATACCTGCCCAACC 1074
Db 589 -----GluGlySerArgTrpValProPr 596
Qy 1075 AACGAATTTTCAGGACTTCCAGATTTCAGTCAGATAGGGCGCTTCCCTCATCATCTC 1134
Db 596 o-----AspSerAlaCysSerSerCysValCy 605
Qy 1135 CTGGAGGTGGTGAGGCCACATCGGC-----TGCCTG-----CTTGCTCAGGG 1179
Db 605 sHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 624
Qy 1180 AGAAGAGCCGAGAAAGAACACTCTCCTCAAGTGGCTGGCAGAGCAGATGATCTCGGAGGA 1239
Db 624 oArgGlnGlyProHisAspCysCysProGlnCys----- 635
Qy 1240 ATTTGGAAGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGATGAAAT 1299
Db 636 -----SerAspCysGluHisGluGlyArgLysTy 645
Qy 1300 GGAG-----TAGAGTATAAGTGTGAATGCATGTTGATTTGTC 1338
Db 645 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 662
Qy 1339 TTAGTCTAGAAATCTAGTTTAGAAAGGATGTTTAGGGGAACATGAGGCTGGCTCTGCA 1398
Db 663 -----CysGI 664
Qy 1399 GCAACAACCGGCTCCCTGTCATCCCTGGGCCAGGAGTTTACTCAGAGCTCTCTGAAG 1458
Db 664 uProGlnProGluGlyProProSerLeu-----Ar 674
Qy 1459 ATGT---GGCAACCCATGCCCTTTTCTGAGGA-----GGTGC 1494
Db 674 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 694
Qy 1495 ATGCCCTCAGCATTTG----- 1510
Db 694 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 714
Qy 1511 -----TTGTCTGGCCCA-----GAGGAGAGAGC 1533
Db 714 uLeuGlySerGluLeuAlaProProAspProCysTyThrCysGlnCysGlnAspLeuTh 734

Qy 1534 TTGGTTCCCAT----- 1546
Db 734 rTriLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 754
Qy 1547 -----GTCCTGGGAGAGTGTCTGCAGGGCGGCGGAGGCGAGAG 1584
Db 754 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArg 774
Qy 1585 CAGCCCTCGGGAGAGCTCACTCTGGTCGACTTCTCTCTCAGAGAATGT----- 1634
Db 774 rg---ValAlaAspGlySerTrpArgAsp---ProSerAsnAlaCysIleAlaCyst 792
Qy 1635 --TGCTCTGGAGG-----TGCTCTGTCATGAAACCCCTAAT 1668
Db 792 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 807
Qy 1669 GGTTCTCTGTTGTTTTCCTTCAAAATTTAGAAATAAGTCTCCGGATGGGCTGTGTGAT 1728
Db 808 ----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-SerCysCysGlu 825
Qy 1729 ACCATTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTCCTGTAGCTAGATATT 1788
Db 826 Arg-----CysGlnAlaProThr 831
Qy 1789 TCCCA-----GAGACACGCGAACTGTCTCTTCTTAAGCCCGGAGAGCGC 1839
Db 832 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 846
Qy 1840 AGCAATGGGCGCTCGCAGGCCAGGCTTGACACAGCATGCTTGTAGTTAGGACTTAAA 1899
Db 847 -----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 862
Qy 1900 ATTATCCAGTTCTCTGCTGTTCTTCTACTTGAATGTGGA 1938
Db 863 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 875

RESULT 11
US-10-189-971-24
; Sequence 24, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 1342
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-24

Alignment Scores:
Pred. No.: 0.000107 Length: 1342
Score: 157.50 Matches: 172
Percent Similarity: 27.78% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 231
Query Match: 4.12% Indels: 373
DB: 15 Gaps: 47

US-09-743-237-3 (1-2134) x US-10-189-971-24 (1-1342)

Qy 14 GCCGAACCTCGCGGGGGCGGACAGCGGGGGAGTCTCTCGGGGAGTACCCGGGAT 73

Db 850 -----GluGlySerArgTrrpValProPr 857
QY 1075 AACGAAATTTTCAGGACTTCCAAGATTCAAGTACAGATAGGCGGCTTCTCTCATGATCTC 1134
Db 857 O-----AspSerAlaCysSerSerCysValcy 866
QY 1135 CTGGAGGTGGTGGAGGCCACATCGCC-----TGCCTG-----CTTGCTCAGGG 1179
Db 866 sHisGluGlyVal-----ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 885
QY 1180 AGAAGAGCCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGATCCTGGAGA 1239
Db 885 oArgGlnGlyProHisAspCysCysProGlnCys 896
QY 1240 ATTTGGAAGTGTCTATCACAGATTCTCCACACTGAGTTAAATCTAAGGATTGAAAT 1299
Db 897 -----SerAspCysGluHisGluGlyArgLysTy 906
QY 1300 GGAG-----TAGAGTATAAGTGTGAATCATGTTGATTTTCTC 1338
Db 906 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 923
QY 1339 TTAGTCTAGAATCTCTAGTTAGAAAGGATGTTTAGGGGACATGAGCTGGCTCTCA 1398
Db 924 -----CysG1 925
QY 1399 GCAACAACAGGCTCCCTGCATCCCTGGGCCAGGAGTTTACTCAGAGCTCTCTGAAG 1458
Db 925 uProGlnProGluGlyProProSerLeu-----Ar 935
QY 1459 ATGT---GGCAACCATCCCTTTCTGAGGA-----GGTGC 1494
Db 935 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 955
QY 1495 ATGGCTCAGCATTTGT----- 1510
Db 955 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 975
QY 1511 -----TTGCTCTGGCCA-----GAGGACAGAGC 1533
Db 975 uLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGlnAspLeuTh 995
QY 1534 TTGGTTCCCATTA----- 1546
Db 995 rTriLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 1015
QY 1547 -----GTCTGGGAGAGTCTCTGAGCGCGCGGAGGAGCAG 1584
Db 1015 rProProGlySerCysCysProValCysArgGluCysValVal -GluAlaGluGlyArgA 1035
QY 1585 CAGGCCCTCGGAGAGCTCACTCTGCTGCTCTCTCTCAGACAATGT----- 1634
Db 1035 rg---ValAlaAspGlyGluSerTrpArgasp---ProSerAsnAlaCysIleAlaCysT 1053
QY 1635 --TGCTCTGGAGGC-----TGCTCTGCATCAAAACCCCTAAT 1668
Db 1053 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 1068
QY 1669 GGTTCCTGTTGTTTTCAATATTATTAGAATAAGTTCCTCGGATGGCTGTTGTGAT 1728
Db 1069 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp -SerCysCysGlu 1086
QY 1729 ACCACTTAAATCTCTAGAACTACTGAACACCTAAAGATTTTCTGTAGCGCTAGATATT 1788
Db 1087 Arg-----CysGlnAlaProThr 1092
QY 1789 TCCCA-----GAGACAGCGCACTGTGATCTTTTCTTAAGCCCCCGGGAGACGC 1839
Db 1093 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 1107
QY 1840 AGGCAATGGGCTCGCAGGCGGCTTGACACCATGCTTGTAGTTAGAGGACTTAA 1899
Db 1108 -----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 1123

QY 1900 ATTATCCAGTTTCTCTGTGTCTTACTTGAATTGTGGA 1938
Db 1124 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1136

RESULT 13

US-10-189-971-12
; Sequence 12, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1570
; TYPE: PR1
; ORGANISM: homo sapiens
US-10-189-971-12

Alignment Scores:

Pred. No.: 0.000111 Length: 1570
Score: 157.50 Matches: 172
Percent Similarity: 27.78% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 231
Query Match: 4.12% Indels: 373
DB: 15 Gaps: 47

US-09-743-237-3 (1-2134) x US-10-189-971-12 (1-1570)

QY 14 GCGAAGCTCGCGGGGGCGGACGAGCGGGGAGCTCTCTGGGGAGTACCCCGGAT 73
Db 541 GlyCysAlaPheGlyGlySerGlyAlaAspPheProHisProSerAsp 560
QY 74 CCCAG-----AGCTCAGCGCTGGAGGACCTCGCGCTCTCCAGG 114
Db 561 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 580
QY 115 CCCCCGACGCGCGCCG-----CCTGCAAGCTGCGACTTCTGTCTCGCTGC 159
Db 581 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 598
QY 160 TACCGCGCAGCGGAGCGCGGGTGTTCCTCCCTGGGGCG----- 201
Db 599 ProAlaAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 618
QY 202 -----CCTGGGTCTCGGAAGAGCCTCCACCGCGGCTCGCATGATC 246
Db 619 GluTyrPheSerProProGlyAsp-----ProCysArgArgCysLeu 632
QY 247 CCAGTTGAAATCAAGTAAGCAGGTGGTACTACTACAAGTAATAATCCGGAAGAACAC 306
Db 633 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 644
QY 307 TTGCGAGATCTTCTTGTCTCAGGAATCTGTTGCAAGTTCCTCATGTCCTCCAGGAAC 366
Db 645 CysProProAlaProCysAlaHis-----ProArgGlnGlyProCysCysProSerCysAsp 663
QY 367 GGATGCCCTCCTG-----CTGTTCCTTAAGAAAGATTTC---CAACCCCAATG 411
Db 664 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAl 683
QY 412 GATATGCCAATTTGAAAGGGGCGCACAAATGTCTATAGAAATTTCTAGAACAGAGA 471

Db 1051 gCysHisArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 1071
Qy 1495 ATGCCCTGAGCATCTCT----- 1510
Db 1071 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 1091
Qy 1511 -----TTGCTGCGCCCA-----GAGGAGAGAGC 1533
Db 1091 uLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGlnAspLeuTh 1111
Qy 1534 TTGGTTCCCATATA----- 1546
Db 1111 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 1131
Qy 1547 -----GTCCTGGGAGAGTGTCTGAGGCGCGCGGAGGAGCAGAG 1584
Db 1131 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA 1151
Qy 1585 CAGGCCCTGGGAGAGCTACTCTGGTGGACTCTCTCTCAGAGAATGT----- 1634
Db 1151 rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCyst 1169
Qy 1635 --TCCTCTGGAGGC-----TGCTCTGCATGAACCCCTAAT 1668
Db 1169 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 1184
Qy 1669 GGTTCCTGTGTGTTTTCAAATATTATAGAAATAAGTCTCCGGATGGGCTGTGTGAT 1728
Db 1185 -----LeuSerCysProHisGlyTrpAlaIysValProGlnAlaAsp-SerCysCysGlu 1202
Qy 1729 ACCACTTAAATCTCTAGAGAAGTACTGAACACCTAAAGATTTCTGTAGCGTAGATATT 1788
Db 1203 Arg-----CysGlnAlaProThr 1208
Qy 1789 TCCCCA-----GAGACAGCGGAAGTGTCTCTTCTTAAGGCCCGCGGAGAGCG 1839
Db 1209 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 1223
Qy 1840 AGCAATGGGCGCTCGCAGCGCAGGCTTGACACAGCATGCTTGAGTTAGAGGACTTAAA 1899
Db 1224 -----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 1239
Qy 1900 ATTATCCAGTTCTCTGTGTTTCTACTTGAATTTGTGGA 1938
Db 1240 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1252

RESULT 15
US-10-213-509-5
; Sequence 5, Application US/10213509
; Publication No. US20030054485A1
; GENERAL INFORMATION:
; APPLICANT: Weiss, Joseph
; APPLICANT: Scott, Matthew
; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES
; FILE REFERENCE: STAN-232
; CURRENT APPLICATION NUMBER: US/10/213,509
; PRIOR FILING DATE: 2002-08-06
; PRIOR FILING DATE: 60/311,720
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4123
; TYPE: PRN
; ORGANISM: H. sapiens
US-10-213-509-5

Alignment Scores:
Pred. No.: 0.00802 Length: 4123
Score: 137.00 Matches: 153
Percent Similarity: 27.03% Conservative: 50
Best Local Similarity: 20.37% Mismatches: 208
Query Match: 3.58% Indels: 340

DB: 15 Gaps: 44
US-09-743-237-3 (1-2134) x US-10-213-509-5 (1-4123)
Qy 7 GGGTCAAGCGCAAGCTCGCGGGG-----GCGACAGCGACGCGGGGAGC 51
Db 2820 GlyAspArgGlnGluLeuGlnGlyCysHisThrValCysGlyThrGlyIleAlaGlySer 2839
Qy 52 TCCTCGGGAGTACCCCGGATCCAGAGCTCA----- 84
Db 2840 LeuGlyAlaGlyValProProSerSerSerGlnPheCysThrLeuArgThrHisGlyMet 2859
Qy 85 -----GCGGCG 90
Db 2860 GlyProThrAspHisSerThrTrpGlyIleGluValPheGlyTrpThrProTrpThrSer 2879
Qy 91 TGGAGGAGC---TCGCGCTCTCTGC-----AGGCCCGC 120
Db 2880 TrpSerSerCysSerGlnSerCysLeuAlaProGlyGlyProGlyTrpArgSerArg 2899
Qy 121 AGCCGC----- 126
Db 2900 SerArgLeuCysProSerProGlyAspSerSerCysProGlyAspAlaThrGlnGluGlu 2919
Qy 127 -----CGCGCTCAACGTCGACT-----TCCTGTCTCT--- 153
Db 2920 ProCysSerProProIleGluCysThrGlyPheCysAlaProGlyCysThrCysProPro 2939
Qy 154 -----CGCTGCTACCGCGCAGCCGCGCGGGGTG--- 186
Db 2940 GlyLeuPheLeuHisAsnAlaSerCysLeuProArgSerGlnCys---ProCysGlnLeu 2958
Qy 187 -----TTTGCCCTCGCGCGCTCTGGG-----TCCTGGAAGGAGCGCTCC 225
Db 2959 HisGlyGlnLeuTyrAlaSerGlyAlaMetAlaArgLeuAspSerCysAsnAsnCysThr 2978
Qy 226 CACCCGCGCTCCGATG-----ATCCAGTTGAATAAACAAGTAAAGC. 267
Db 2979 CysValSerGlyLysMetAlaCysThrSerGluArgCysProValAlaCysGlyTrpSer 2998
Qy 268 AGGTGGTACTACTACAAGTAATAATCCGGAAGAGCAACTTTGCAGAACTCTCTGTCTCA 327
Db 2999 ProTrpThrLeuTrp-----SerLeuCysSer 3007
Qy 328 GGAATCCTGT----- 337
Db 3008 Cys-SerCysAsnValGlyIleArgArgPheArgAlaGlyThrAlaProProAlaAl 3027
Qy 338 -----TGCAAGTTCCCATGTCCTCCAGGAAGTACAGAGATGCTCTGCTG 381
Db 3027 aPheGlyGlyAlaGluCysGlnGlyPro-----ThrMetGluAlaGluPheCy 3043
Qy 382 TTCTCTTAAAGAAAGATTCCCAACCAATGGTGATATGCCAATTGAAGGGGACACAAAT 441
Db 3043 sSerLeuArgProCysProGlyProValProGlyMetCysProArgAspLysGlnTrpLe 3063
Qy 442 GCTATGTATAGACAAT---TCTAGAACAAGAACTAAAGCACTCCATTGGTGGTCTCA 498
Db 3063 uAspCysAlaGlnGlyProAlaSerCysAlaGluLeuSerAla-----ProAr 3079
Qy 499 GTATCAAGATCAAAATAATTATCTACAGTACAGATGTCCTTAAACCAATGACTGCTTAGT 558
Db 3079 gGlyThrAsnGlnThrCysHisProGlyCysHisCysProSerGlyMetLeuLeuLeu 3099
Qy 559 AGGG---AGATTTTTCGACGATCAACAAAAATTAATCTCATTTACACAACAACTTGAGG 615
Db 3099 lSerProArgGlyHisProGlyProLeuGlyAlaSerVal-----GlnProProVa 3116
Qy 616 AGCCTTACCATCGGTAGTCAACGGGTCTGCTTTCCCTCGGGATCA----- 661
Db 3116 lAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAlaGlyGlyTrpGlyProTr 3136
Qy 562 -----AC 663


```
Db 3136 pGlyProTrpSerHisCysSerArgSerCysGlyGlyGlyLeuArgSerArgThrArgAl 3156
QY 664 TCTTCCAGGACCAACAAAATAACTTTGGCTGGGTACTGTGAC---TCGTTTGGCAGTGG 720
Db 3156 aCysAspGlnProProGlnGlyLeuGlyAspTyrCysGlyGlyProArgAlaGlnGl 3176
QY 721 GCACCTTTTTCG-----AACAACTGCAAT----- 742
Db 3176 yGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGluGlyAlaGl 3196
QY 743 -----TGTAATAATTTGTTGCAACAACTTGCATCATGATATTGA 780
Db 3196 uTyrSerProCysGlyProProCysProArgSerCysAspAspLeuValHis----- 3213
QY 781 ACGGTTTAAGCCATTAGGCATGCTTGGTAGA---AATCCAGAAGCTTTCCAGCCAAA 837
Db 3214 -----CysValTrpArgCysGlnProGlyCysTyrCysPro-- 3225
QY 838 AATTGGGAAGGCCAAATTGGCAAT-----GTCAAGCCGCCAGCACAA 879
Db 3226 -----ProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnProGlyHis-- 3242
QY 880 CAAGGGTGCACACTGC----- 895
Db 3243 -----CysSerCysLeuAspLeuLeuThrGlyGlnArgHisProGlyAlaArgLe 3260
QY 896 ---AGGAGTCCAGCTCCGTCAGAAATTAAGTCCAGTGCCTATGAGGCCCAAAATTATGTG 951
Db 3260 uAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGluGlyArgLeuAsnCy 3278
QY 952 TTCCTCTATTGCAATTCATTTGTCGAAAAAATTATGAAGAAAGCCCAAGAACGAAAGAC 1011
Db 3278 sThrAspLeu---ProCysProAspCys----- 3286
QY 1012 ACTAATAGCATGCCAACTACATGCAGACTGGAGTTTGAAGGACGCCATTACCTGCC 1071
Db 3287 -----GlyGlyGlyGlnSerLeuHis----- 3293
QY 1072 ACCAAGAAATTTTCAGGACTTCCAGATTCCAGATTCAGTCACGATAGCGGCTTCTCTCA---TG 1128
Db 3294 -ProCysGly---GlnProCysProArgSerCysGlnAspLeuSerProGlySerValCy 3312
QY 1129 CATCTCCCTGGGAGTGTG---GTGGAGGCCACATGCGCCTGCTTGTCTGCTCAGGGAGAAGA 1185
Db 3312 sGlnProGlySerValGlyCysGlnProThrCysGlyCysProLeuGlyGln----- 3329
QY 1186 GCGCGAGAAAGAACTGCTCCAGTGCCTGGCAGCAGCATGATCCTGGAGGAATTTGG 1245
Db 3330 -----LeuSerGlnAspGl 3334
QY 1246 AAGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGGATTGAAAATGGAGTA 1305
Db 3334 yLeuCysValPro----- 3338
QY 1306 GAGTATAAAGTGTGAATGCATGTTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAA 1365
Db 3339 -ProAlaHisCysArgCys----- 3344
QY 1366 GGATCTTTAGGGGAACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCTGCA---TC 1422
Db 3345 -----GlnTyrGlnProGlyAlaMetAlaProSe 3354
QY 1423 CCTGGGCCAGGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAAC-----CCATGCC 1476
Db 3354 rPheValProSerThrCysValAlaIleGlyLeuGlnCysGlnGluValProAspCysPr 3374
QY 1477 CCTTTTCTGAGGAGGTGTCGCTGAGCATTGTTTGTCTGGCCAGAGGAGAGAGCTTG 1536
Db 3374 o-----AspProGlyValTrpSerTrp-----GlyProTrpGluAspCysSe 3389
QY 1537 GGTTCCTCCATAGTCTCTGGGAGAGTGTCTGCAG----- 1567
```

```
Db 3389 rValSerCysGlyGlyGlyGluGlnLeuArgSerArgArgCysAlaArgProProCysPr 3409
QY 1568 -GGCGGCGAGGGGAGAGCAGGCGCTGCGGAGAGCTCACTGTGGTCGACTCTTCTCTCA 1626
Db 3409 oGlyProAlaArgGlnSerArgThrCys-----SerThrGl 3421
QY 1627 GAGAATGTTGCTCTGAGGCTGCTCTGCA 1655
Db 3421 n-ValCysArgGluAlaGlyCysProAla 3430
```

Search completed: July 24, 2003, 14:44:31
Job time : 146.043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:50:19 ; Search time 32.9246 Seconds

(without alignments)

5484.741 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aatcggtggaaggcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table:

BLOSUM62

Xgapop 10.0, Xgapext 0.5

Fgapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp

-Q/cgn2_1/USPTO_spool/US09743237/runat_24072003_113106.17736/app_query.fasta_1.4750

-DB-issued_Patents_AA -Qfmt=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPECL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi

-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09743237@cgn_1.100 @runat_24072003_113106.17736 -NCPU=6 -ICPU=3

-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	3.3	141	US-09-252-991A-25325	Sequence 25325, A
2	125	3.3	483	US-09-252-991A-19958	Sequence 19958, A
3	124	3.2	739	US-08-444-818-148	Sequence 148, App
4	124	3.2	2995	US-08-444-818-138	Sequence 138, App
5	118.5	3.1	320	US-09-252-991A-24634	Sequence 24634, A
6	117	3.1	206	US-09-252-991A-31481	Sequence 31481, A
7	117	3.1	231	US-09-252-991A-29962	Sequence 29962, A
8	117	3.1	1021	US-07-910-760-12	Sequence 12, Appl
9	117	3.1	1021	US-08-440-519-12	Sequence 12, Appl
10	117	3.1	1021	US-08-440-549-12	Sequence 12, Appl
11	117	3.1	2261	US-08-444-818-66	Sequence 66, Appl
12	117	3.1	2436	US-08-444-818-75	Sequence 75, Appl

13	117	3.1	2772	3	US-08-444-818-89	Sequence 89, Appl
14	117	3.1	2894	2	US-08-466-975A-23	Sequence 23, Appl
15	117	3.1	2894	2	US-08-391-671A-23	Sequence 23, Appl
16	117	3.1	2894	3	US-08-467-902A-23	Sequence 23, Appl
17	117	3.1	2894	3	US-09-275-265-23	Sequence 23, Appl
18	117	3.1	2894	4	US-09-941-611-23	Sequence 23, Appl
19	117	3.1	3011	1	US-08-440-103-36	Sequence 36, Appl
20	117	3.1	3011	1	US-08-440-542-36	Sequence 36, Appl
21	117	3.1	3011	1	US-07-910-760-10	Sequence 10, Appl
22	117	3.1	3011	1	US-08-440-519-10	Sequence 10, Appl
23	117	3.1	3011	1	US-08-231-368-36	Sequence 36, Appl
24	117	3.1	3011	1	US-08-440-210-36	Sequence 36, Appl
25	117	3.1	3011	3	US-09-388-874-2	Sequence 2, Appl
26	117	3.1	3011	4	US-09-046-604-36	Sequence 36, Appl
27	117	3.1	3011	4	US-08-440-549-10	Sequence 10, Appl
28	117	3.1	3011	4	US-08-850-328-1	Sequence 1, Appl
29	117	3.1	3011	4	US-09-916-359-2	Sequence 2, Appl
30	116	3.0	1692	3	US-09-263-933-4	Sequence 4, Appl
31	116	3.0	1692	3	US-09-263-933-11	Sequence 11, Appl
32	116	3.0	1692	3	US-09-263-933-18	Sequence 18, Appl
33	116	3.0	2013	1	US-08-324-977-12	Sequence 12, Appl
34	116	3.0	2013	2	US-08-384-616-12	Sequence 12, Appl
35	116	3.0	2013	2	US-08-904-686A-12	Sequence 12, Appl
36	116	3.0	2201	4	US-08-952-981A-2	Sequence 2, Appl
37	116	3.1	2294	4	US-09-252-991A-17231	Sequence 2, Appl
38	116	3.1	2307	3	US-09-263-933-2	Sequence 9, Appl
39	116	3.0	2307	3	US-09-263-933-9	Sequence 16, Appl
40	116	3.0	2307	3	US-09-263-933-16	Sequence 32, Appl
41	116	3.0	2620	1	US-08-324-977-32	Sequence 32, Appl
42	116	3.0	2620	2	US-08-384-616-32	Sequence 32, Appl
43	116	3.0	2620	2	US-08-904-686A-32	Sequence 32, Appl
44	116	3.0	2620	3	US-09-315-850-32	Sequence 32, Appl
45	116	3.0	2620	3	US-09-315-850-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1

US-09-252-991A-25325
; Sequence 25325, Application US/09252991A
; Patent No. 6531795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25325
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-25325

Alignment Scores:
Pred. No.: 0.000685 Length: 141
Score: 126.00 Matches: 34
Percent Similarity: 51.90% Conservative: 7
Best Local Similarity: 43.04% Mismatches: 22
Query Match: 3.29% Indels: 16
DB: 4 Gaps: 5

US-09-743-237-3 (1-2134) x US-09-252-991A-25325 (1-141)

QY 10 TCAAGCGGAAGCTCGCGGGGGCGACAGCGCGGGAGCTCTCGGGGAGTACCCG 69

Db 43 SerProArgAlaCysAsnGlyAlaSerThrThrThrArgSerSer---AlaThrPro 61

Sequence 148, Application US/08444818
 Patent No. 6150087
 GENERAL INFORMATION:
 APPLICANT: Chien, David Y.
 APPLICANT: Rutter, William J.
 TITLE OF INVENTION: NANV Diagnostics and Vaccines
 NUMBER OF SEQUENCES: 777
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Chiron Corporation
 STREET: 4560 Horton Street
 CITY: Emeryville
 STATE: CA
 COUNTRY: USA
 ZIP: 94608-2916
 COMPUTER READABLE FORM:
 MEDIUM TYPE: IBM PC compatible
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/444,818
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/403,590
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Harbin, Alisa A.
 REGISTRATION NUMBER: 33,895
 REFERENCE/DOCKET NUMBER: 0110.002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508)359-3876
 TELEFAX: (508)359-3885
 INFORMATION FOR SEQ ID NO: 148:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 739 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-444-818-148

Alignment Scores:
 Pred. No.: 0.0023 Length: 739
 Score: 124.00 Matches: 91
 Percent Similarity: 35.32% Conservative: 45
 Best Local Similarity: 23.64% Mismatches: 140
 Query Match: 3.24% Indels: 109
 DB: 3 Gaps: 22

US-09-743-237-3 (1-2134) x US-08-444-818-148 (1-739)

QY 641 TCTGCTTTCCCC---TCGGATCAACTCTCCAGGACCACCAAAATAACTTTGGCTGG 697
 Db 233 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 252
 QY 698 TACTGTGACTGCTTTCCAGTGGGACTTTTGCACAACTGCAATGTAATAATTCTTGC 757
 Db 253 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 267
 QY 758 AACAACTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
 Db 268 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 287
 QY 797 AAGGCATGCTGGTGAATAATCCAGAGCTTTCCAGCCCAAAATTTGGGAGGCCAATTG 856
 Db 288 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThr 300
 QY 857 GGCATGTCAAGCCCAAGCAGCACACAAA-----GGGTGCAACTGCAGGAGGTCAGGCTGC 910
 Db 301 GlyArgGlyLysProGlyIleAsnArgPheValAlaProGlyGluArgProSerGlyMet 320
 QY 911 CTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAAAATTATGTGTTCTTATTTCG 964

Db 321 PheAspSerSerValLeuCysGluCysTyrAspAla----- 332
 QY 965 AAATGCATTGGTTGCAAAATTTATGAAGAAAGCCAGACACACTAATGACCATG 1024
 Db 333 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArgLeu 348
 QY 1025 CCAAACTACATGCAGACTCGAGTTTG-----GAA 1054
 Db 349 ArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGlu 368
 QY 1055 GGC-----AGCCATTACTGCCCAACACAGAAATTTTCA 1087
 Db 369 GlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSer 388
 QY 1088 GGA-----CTTCCAGATTTCAGTC-----ACGATAGCGGCTTCTCTCATGC 1129
 Db 389 GlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAl 408
 QY 1130 ATCTCTCTGGGAGGTGG-----TGGAGGCCACATGGCGCTGCTGCTGTTGCTCAGGGA 1180
 Db 408 aProProSerTrpAspGlnMetTrpLys-----CysLeuIleArgLe 423
 QY 1181 GAAGAGCCGCGAAGAACACTGCTCCAAGTGCCTGGCAGCAGAGCATGATCCTGGAGGAA 1240
 Db 423 uLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 437
 QY 1241 TTGGGAAGTGTCTTATCACAGATTCTCCACACTGACT----- 1277
 Db 438 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMe 455
 QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGAGTATAAAGTGT 1318
 Db 455 tThrCysMetSerAlaAspLeuValValThrSerThrTrpValLeuValGlyGlyVa 475
 QY 1319 GAATGCA-----TGTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTACAAGGATGTT 1372
 Db 475 lLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal----- 492
 QY 1373 TAGGGGAACATGAGGCTGCTCTGCAGCAACACAGGCTCCCTCCATCCCTGGGCCCA 1432
 Db 493 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArgG 508
 QY 1433 GGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCTTTTCTGA 1486
 Db 508 uValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIleG 528
 QY 1487 GGAGTGCATGGCCTGAGCATTTGTTGCTGCGCCAGAGAGAGAGCTGGGTGCCATA 1546
 Db 528 uGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu----- 544
 QY 1547 GTCTGGGAGAGTGTCTGCGAGCGCGGCGGAGGAGGAGGCGCTGCGGAGAGCTCACT 1606
 Db 545 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnThrAs 562
 QY 1607 CTGCTCGACTCTT 1619
 Db 562 nTrpGlnLysLeu 566

RESULT 4

US-08-444-818-138
 Sequence 138, Application US/08444818
 Patent No. 6150087
 GENERAL INFORMATION:
 APPLICANT: Chien, David Y.
 APPLICANT: Rutter, William J.
 TITLE OF INVENTION: NANV Diagnostics and Vaccines
 NUMBER OF SEQUENCES: 777
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Chiron Corporation
 STREET: 4560 Horton Street
 CITY: Emeryville
 STATE: CA
 COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 2995 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-138

Alignment Scores:
Pred. No.: 0.00412 Length: 2995
Score: 124.00 Matches: 91
Percent Similarity: 35.32% Conservative: 45
Best Local Similarity: 23.64% Mismatches: 140
Query Match: 3.24% Indels: 109
DB: 3 Gaps: 22

US-09-743-237-3 (1-2134) x US-08-444-818-138 (1-2995)

QY 641 TCTGCTTCCCTCC---TCGGGATCAACTCTTCAGGACCCACCAAAATAACTTTGGCTGGG 697
Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
QY 698 TACTGTGACTGTTGCCAGTGGGACTTTGCAACAACACTGCAATTGTAATAATTTGGC 757
Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
QY 797 AAGCATCTCTGGTGGAGAAATCCAGAGCTTTCAGCCCAAAATTTGGAGGGCCAAATTG 856
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThr 1491
QY 857 GCATGTCACAGCCCAACACACAA-----GGGTCAACTGCAGAGGTTCAGGCTGC 910
Db 1492 GlyArgGlyLysProGlyIleAsnArgPheValAlaProGlyGluArgProSerGlyMet 1511
QY 911 CTGAAGAT-----TACTGCGAGTGTATGAGGCCCAAAATATGCTGTTCTCTATTGTC 964
Db 1512 PheAspSerSerValLeuCysGluCysThrAspAla----- 1523
QY 965 AAATGCATTTGGTGCACAAATATGAGAACCCAGAACACACACTAATGAGCATG 1024
Db 1524 -----GlyCysAlaTrpThrGluLeuThrPro---AlaGluThrThrValArgLeu 1539
QY 1025 CCAAACTACATCCACACTGGAGTTTG-----GAA 1054
Db 1540 ArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGlu 1559
QY 1055 GGC-----AGCCATTACCTGCCACCAACACAAATTTTCA 1087
Db 1560 GlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSer 1579

QY 1088 GGA-----CTTCCAAGATTCAAGTC-----ACGATAGCGCGCTTCTCTCATGC 1129
Db 1580 GlyGluAsnLeuProThrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAl 1599
QY 1130 ATCTCTCTGGAGGTGG-----TGGAGGCCACATGCGCTGCTGCTGCTCAGGGA 1180
Db 1599 aProProSerTrpAspGlnMetTrpLys-----CysLeuIleArgLe 1614
QY 1181 GAAGAGCGCGAGAAACACACTGCTCCAAGTCCCTGGCAGACGACATGCTCGGAGAA 1240
Db 1614 ulysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
QY 1241 TTTGGAAGGTGCTTATCATCAGATTCTCCACACTGAGT----- 1277
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMe 1646
QY 1278 -----TTAAATCTAAGGATTTGAAATGGAGTAGTATAGTATAAAGTGT 1318
Db 1646 tThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyVa 1666
QY 1319 GAATGCA-----TGTTCATTTTGTCTAGTCTAGAAATCTCTAGTTTAGAAAGCATGTT 1372
Db 1666 lLeuAlaAlaLeuAlaIleThrCysLeuSerThrGlyCysValValIleVal----- 1683
QY 1373 TAGGGAAACATGAGGCTGCTCTGCAGCAACCAACAGGCTCCCTGTCATCCCTGGGCCCA 1432
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArgGl 1699
QY 1433 GGGAGTTTACTCAGAGCTCTCT-----GAGATGTGGCAACCCATGCCCTTTTCTGA 1486
Db 1699 uValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIleGl 1719
QY 1487 GGAGTGCATGCGCTGAGCATTTGTTGCTGGCCCGAGGAGAGAGCTTGGGTTCCCATATA 1546
Db 1719 uGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu----- 1735
QY 1547 GTCTTGGGAGAGTGTCTGCGAGGCGCGGAGGCGAGAGCGCCCTGCGGAGAGCTCACT 1606
Db 1736 ---LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnThrAs 1753
QY 1607 CTGCTCGACTCTT 1619
Db 1753 ntrpGlnLysLeu 1757

RESULT 5

US-09-252-991A-24634
Sequence 24634, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS.
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24634
LENGTH: 320
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24634

Alignment Scores:
Pred. No.: 0.00532 Length: 320
Score: 118.50 Matches: 71
Percent Similarity: 32.90% Conservative: 30
Best Local Similarity: 23.13% Mismatches: 140
Query Match: 3.10% Indels: 66

QY 1052 GAAGGC-----AGCCATTACCTGCCACCAAGAAATTT 1084
Db 527 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 546
QY 1085 TCAGGA-----CTTCAAGATTCAGTC-----ACGATAGCGCGCTTCCTCA 1126
Db 547 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 566
QY 1127 TGCATCTCTCGGAGGTGG-----TGGAGGCCACATCGCGCTCCCTGCTTGCCTCAG 1177
Db 566 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 581
QY 1178 GGAGAGAGCGCCGAGAAGAACTGCTCCAAAGTGCCTGCAGAGAGATGATCCTGGAG 1237
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 596
QY 1238 GAATTTGAGAGTGTCTATCACAGATTCACACACTCAGT----- 1277
Db 597 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 613
QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGAGTATAAAG 1315
Db 613 emetThrCysmetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 633
QY 1316 TGTGAATGCA-----TGTGATTTCTCTAGTCTAGAAATCTCTAGTTTAGAAAGGAT 1369
Db 633 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 651
QY 1370 GTTTAGGGGAACATGAGCTGCTGTCAGACACACACAGGCTCCCTGCTGATCCCTGGGC 1429
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 666
QY 1430 CCAGGGAGTTTACTTCAGAGCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC 1483
Db 666 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 686
QY 1484 TGAGGAGGTGCATGGCTGAGCATGTTTCTGCGCCAGAGAGAGAGCTTGGGTGCC 1543
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 703
QY 1544 ATAGTCTGGAGAGTCTGTCAGGCGCGGAGGCGAGAGAGGCTCCGCGAGAGCTC 1603
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla-----ProAlaValGlnTh 720
QY 1604 ACTCTGCTGCACTCTT 1619
Db 720 rAsnTrpGlnLysLeu 725

RESULT 9
US-08-440-519-12
; Sequence 12, Application US/08440519
; Patent No. 5712087
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,519

FILING DATE: 12-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/910,760
FILING DATE: 07-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blackburn Esq., Robert P.
REGISTRATION NUMBER: 30,447
REFERENCE/DOCKET NUMBER: 0101.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2702
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-519-12

Alignment Scores:
Pred. No.: 0.0124 Length: 1021
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 23 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-440-519-12 (1-1021)

QY 641 TCTGCTTTCCCC---TCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTTGGCTGGG 697
Db 392 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 411
QY 698 TACTGTGACGCTTGGCCAGTGGGACTTTTGCACCACTGCAATTTGTAATAATGTTGC 757
Db 412 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 426
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrLeuPro 446
QY 797 AAGGCATGCTTGGTAGAAATCCAGAGCTTCCAGCCCAAAATTTGGAGGGCCCAATTG 856
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAAGCCCCACACACAAAGGTGCTCACTGCAGGAGGTGAGGC 907
Db 465 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAATATATGTGTTCTCTATT 961
Db 479 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 491
QY 962 TGCAATGTCATTTGTTGCAAAATTTATGAAGAAAGCCCAAGAACAGACACTAATGAGC 1021
Db 492 -----GlyCysAlaTrpTyrGluLeuThrPro-----AlaGluThrThrValArg 506
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTTG----- 1051
Db 507 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
QY 1052 GAAGGC-----AGCCATTACCTGCCACCAACCAAAATTT 1084
Db 527 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 546
QY 1085 TCAGGA-----CTTCAAGATTCAGTC-----ACGATAGCGCGCTTCCTCA 1126
Db 547 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 566
QY 1127 TGCATCTCTCGGAGGTGG-----TGGAGGCCACATCGCGCTCCCTGCTTGCCTCAG 1177
Db 566 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 581

```
QY 1178 GGAGAAGAGCGCGAGAAACACACTCTCCAAAGTCCTGGCAGCAGCAGATGATCCTGGAG 1237
|   |||
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu-----596
QY 1238 GAATTTGGAGGTCTTATCAGATCTCCACACTGACT-----1277
|   |||
Db 597 -----GlyAlaValGlnAsnGluThrLeuThrHisProValThrLysTyrI 613
QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGATATAAAG 1315
|   |||
Db 613 eMetThrCysMetSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyG 633
QY 1316 TGTGAATGCA-----TGTGATTTTCTTACTAGAAATCTCTAGTTTACAAAGGAT 1369
|   |||
Db 633 yValLeuAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----651
QY 1370 GTTTAGGGAACATGAGGTGCTCTGCAGCACAACACAGGCTCCCTGCATCCCTGGGC 1429
|   |||
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleProAspAr 666
QY 1430 CCAGGAGTTTACTCAGACTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC 1483
|   |||
Db 666 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI 586
QY 1484 TGAGGAGGTGCATGGCCTCAGCATTTGTTCTGTGCCCCAGAGGAGCAGCTTGGGTTCCC 1543
|   |||
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu--703
QY 1544 ATAGTCCTGGGAGAGTGTCTGCAGGCGCGGAGGCGCAGAGCCCTGGCAGAGCTC 1603
|   |||
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 720
QY 1604 ACTCTGGTGCAGCTCTT 1619
|   |||
Db 720 rAsnTrpGlnLysLeu 725

RESULT 10
US-08-440-549-12
; Sequence 12, Application US/08440549
; Patent No. 6312889
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigen for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,549
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/910,760
; FILING DATE: 07-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
```

```
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1021 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-549-12
```

```
Alignment Scores:
Pred. No.: 0,0124 Length: 1021
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 4 Gaps: 23
```

US-09-743-237-3 (1-2134) x US-08-440-549-12 (1-1021)

```
QY 641 TCTGCTTTCCCC---TCGGGATCAACTTTCAGGACCACCAAAATAACTTTTGCTGGG 697
|   |||
Db 392 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 411
QY 698 TACTGTGACTGCTTTGCCAGTGGGAGCTTTTGCACAACACTGCAATTGTATAATTGCTTC 757
|   |||
Db 412 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 426
QY 758 AACAACTTG-----CATCATGATATTGACGGTTTAAAGCCATT 796
|   |||
Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrLeuPro 446
QY 797 AAGGCATGCTTTGGTAGAAATCCAGAAGCTTTCCAGCCAAAATTTGGGAAGGCCAAATTG 856
|   |||
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAGGCCCCAGCACACAAGGGTGCACACTGCAGGAGGTGAGGC 907
|   |||
Db 465 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCTGGAAGAAAT-----TACTGCGAGTGTATGAGGCCCAAAATATTGTGTTCTCTATT 961
|   |||
Db 479 MetPheAspSerSerValLeuCysGluCysTyrAspAla-----491
QY 962 TCGAAATGCAATTTGTTGCAAAATATTATGAAGAACCCAGAACAGAACACATAATGAGC 1021
|   |||
Db 492 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg 506
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTTG-----1051
|   |||
Db 507 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
QY 1052 GAAGGC-----AGCCATTACTGCGCACCAACAGAAATTT 1084
|   |||
Db 527 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 546
QY 1085 TCAGGA-----CTTCCAAGATTTCAGTC-----ACATAGCGCGCTTCTCTCA 1126
|   |||
Db 547 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG 566
QY 1127 TGCATCTCTCTGGGAGGTGG-----TCGAGGCCACATGCGCTGCTGCTGCTCAG 1177
|   |||
Db 566 nAlaProProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 581
QY 1178 GGAGAAGAGCGCGAGAAACACACTCTCCAAAGTCCTGGCAGCAGCAGATGATCCTGGAG 1237
|   |||
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu-----596
QY 1238 GAATTTGGAGGTCTTATCAGATCTCCACACTGACT-----1277
|   |||
Db 597 -----GlyAlaValGlnAsnGluThrLeuThrHisProValThrLysTyrI 613
QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGATATAAAG 1315
|   |||
```

Db 613 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 633
QY 1316 TGTGAATGCA-----TGTGATTTTCTTCTAGCTAGAAATCTCTAGTTTAGAAAGGAT 1369
Db 633 yValLeuAlaLeuAlaAlaThrCysLeuSerThrGlyCysValValIleVal----- 651
QY 1370 GTTTAGGGAACATGAGCTGCTCTGCAGCAACAAACAGGCTCCCTGCATCCCTGGGC 1429
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 666
QY 1430 CCAGGAGTTTACTCAGACTCTCT-----GAAGATGTGGCAACCCATCCCTTTTC 1483
Db 666 gGluValLeuThrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 686
QY 1484 TGAGGAGGTGCTGAGCATTTCTTCTGCGCCAGAGGAGAGAGCTTGGGTTC 1543
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 703
QY 1544 ATAGTCTGGGAGAGTGTCTGAGGGGGGGAGGCGAGAGGCTGCGGAGAGCTC 1603
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 720
QY 1604 ACTCTGGTGCAGCTCTT 1619
Db 720 rAsnTrpGlnLysLeu 725

RESULT 11

US-08-444-818-66
; Sequence 66, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2261 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-66
Alignment Scores:
Pred. No.: 0.0176 Length: 2261
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46

Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 3 Gaps: 23
US-09-743-237-3 (1-2134) x US-08-444-818-66 (1-2261)
QY 641 TCTGCTTTCCCC---TCGGATCAACTCTTCCAGGACCACCAAAAAATAACTTTGGCTGGG 697
Db 799 SerValIleProThrSerGlyAspValValValValAlaThrAspAlaLeuMetThrGly 818
QY 698 TACTGTGACTGCTTTGCCAGTGGGACTTTTGCACAACTGCAATGTATAATATTGTTC 757
Db 819 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 833
QY 758 AACCAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 834 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 853
QY 797 AAGGCATGTCTTGGTAGAAATCCAGAACTTCCAGCCAAAAATTTGGGAAGGCCCAATTC 856
Db 854 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 871
QY 857 GGC-----AATGTCAAGCCCCCAGCACACAAGGGTGCACCTCCAGGAGGTACGC 907
Db 872 GlyIleThrArgPheValAlaProGlyGlu-----ArgProSerGly 885
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAAAATTATGTCTTCTTATT 961
Db 886 MetPheAspSerSerValLeuCysGluCysThrAspAla----- 898
QY 962 TGCAAAATGCATTGGTTGCAAAAAATTTATGAAGAAACCCAGACGAAAGACACTAATAGC 1021
Db 899 -----GlyCysAlaThrTyrGluLeuThrPro--AlaGluThrValArg 913
QY 1022 ATCCCAACTACATGCAGACTGGAGTTTG----- 1051
Db 914 LeuArgAlaThrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 933
QY 1052 GAAGGC-----AGCATTACCTGCCCAACCAAGAAATTT 1084
Db 934 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 953
QY 1085 TCAGGA-----CTTCCAAGATTCACTC-----ACGATGCGCGCTTCTCTCA 1126
Db 954 SerGlyGluAsnLeuProTyrLeu-ValAlaThrGlnAlaThrValCysAlaArgAlaG1 973
QY 1127 TGCATCTCTCGGGAGGTGG-----TCGAGGCCACATCGCGCTCTGCTGCTCAG 1177
Db 973 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 988
QY 1178 GGAGAAGAGGCCGCGAAGAAACACTGCTCCAAGTGCCTGCAGACAGATGATCCTGGAG 1237
Db 988 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1003
QY 1238 GAATTTGGAAGGTCTTATCACAGATTCTCCACACTGAGT----- 1277
Db 1004 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1020
QY 1278 -----TTAAATCTAAGGATTGAAATGGAGTAGATATTAAG 1315
Db 1020 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuGlyG1 1040
QY 1316 TGTGAATGCA-----TGTGATTTTCTTCTAGCTAGAAATCTCTAGTTTAGAAAGGAT 1369
Db 1040 yValLeuAlaLeuAlaIleThrCysLeuSerThrGlyCysValValIleVal----- 1058
QY 1370 GTTTAGGGAACATGAGGCTGCTCTGCAGCAACCAAGGCTCCCTGCATCCCTGGGC 1429
Db 1059 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1073
QY 1430 CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATCCCTTTTC 1483
Db 1073 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1093

Qy 1484 TGAGGAGGTGCTGAGCGCTGAGCAATGTTTGTCTGGCCCCAGAGGAGAGAGCTTGGGTTCCTCC 1543
 Db 1093 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1110
 Qy 1544 ATAGTCTCTGGGAGAGTCTCTGCAGCGGGCGGGAGGGCAGAGCAGCGCCCTCGCGAGAGCTC 1603
 Db 1111 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1127
 Qy 1604 ACTCTGGTGCAGCTCTT 1619
 Db 1127 rAsnTrpGlnLysLeu 1132

RESULT 12
 US-08-444-818-75
 : Sequence 75, Application US/08444818
 : Patent No. 6150087
 : GENERAL INFORMATION:
 : APPLICANT: Chien, David Y.
 : APPLICANT: Rutter, William J.
 : TITLE OF INVENTION: NANBV Diagnostics and Vaccines
 : NUMBER OF SEQUENCES: 777
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Chiron Corporation
 : STREET: 4560 Horton Street
 : CITY: Emeryville
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94608-2916
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/444,818
 : FILING DATE:
 : CLASSIFICATION: 424
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US/08/403,590
 : FILING DATE: 14-MAR-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Harbin, Alisa A.
 : REGISTRATION NUMBER: 33,895
 : REFERENCE/DOCKET NUMBER: 0110.002
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (508)359-3876
 : TELEFAX: (508)359-3885
 : INFORMATION FOR SEQ ID NO: 75:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2436 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-444-818-75

Qy	758	AACAAC	TG-----CATCATGATATTGAACGGTTTAAAGCCAA	T	796
Db	1009	ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro			1028
Qy	797	AAGCATGCTTGGTAGAANAATCCAGAGCTTTCAGGCCAA	AAATTTGGGAGAGGCCAA	T	856
Db	1029	GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro			1046
Qy	857	GGC-----AATGTCAAGCCCCAGCACCAACAAGGTTGCAACTGCAGGAGGT	CAGGC	907	
Db	1047	GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly			1060
Qy	908	TGCCTGAAGAAT-----TACTGGAGTGTCTAGAGCCCA	AAATATATGTCTTCTCTATT	961	
Db	1061	MetPheAspSerValLeuGluCysTyrAspAla-----			1073
Qy	962	TGCAAAATGCATTTGGTTCACAAAAATATGAGAAGCCCA	AGAACACACAAATCAATCAGC	1021	
Db	1074	-----GlyCysAlaTTPtyrGluLeuThrPro--AlaGluThrThrValArg			1088
Qy	1022	ATGCCAAACTACATGCAGACTGGAGTTTG-----			1051
Db	1089	LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp			1108
Qy	1052	GAAGCC-----AGCCATTACCTGGCCACCACCAAGAAATTT			1084
Db	1109	GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln			1128
Qy	1085	TCAGGA-----CTTCCAAGATTCACTGTC-----ACGATAGCGCGCTTCTCTCA			1126
Db	1129	SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGln			1148
Qy	1127	TGATCTCTCGGAGGTGG-----TGGAGGCCACATGCCCGCTGCCCTGCTGCAC			1177
Db	1148	nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg			1163
Qy	1178	GGAGAAGCGCCGAGAAAGAACACTGCTCAAGTGCCTGGCAGCAGCAGATGATCCTGGAG			1237
Db	1163	gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu-----			1178
Qy	1238	GAATTTGGAAGGTGCTTATCACAGATTCTCCACACTGAGT-----			1277
Db	1179	-----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle			1195
Qy	1278	-----TTAATCTAAGGATTGAAAATGGAGTAGAGTATATAAG			1315
Db	1195	eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuGlyGln			1215
Qy	1316	TGTGAATGCA-----TGTGTGATTTGTCTTAGTAATAATCTAGTTTGAAGAGGAT			1369
Db	1215	yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal-----			1233
Qy	1370	GTTTAGGGGAACATGAGCTGGCTCTGCCACCAACACAGGCTCCCTGCATCCCTGGGC			1429
Db	1234	-----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArg			1248
Qy	1430	CCAGGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCTTTTC			1483
Db	1248	gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIle			1268
Qy	1484	TGAGGAGGTGCATGGCTGAGCATTTTCTCTGGCCCGACAGAGCAGAGCTTGGTTCCC			1543
Db	1268	eGluGln-----GlyMetMetLeuAlaGluInPheLysGlnLysAlaLeuGlyLeu--			1285
Qy	1544	ATAGTCTCTGGGAGTGCTCTGCAGGGCGGGGAGGCGAGCAGAGCGCTCGCGAGAGCTC			1603
Db	1286	-----LeuGlnThrAlaSerArgGlnAlaGluValIleAla--ProAlaValGlnThr			1302
Qy	1604	ACTCTGCTGCACCTTT	1619		
Db	1302	rAsnTrpGlnLysLeu	1307		

RESULT 13
US-08-444-818-89
; Sequence 89, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-89

Alignment Scores:
Pred. No.: 0.0193 Length: 2772
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 3 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-444-818-89 (1-2772)

QY	641	TCGCTTTCCCC	---TCGGGATCAACTCTTCAGGACACCAAAAATAACTTTGGCTGGG	697
DB	1310	ServAlleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly	1329	
QY	698	TACTGTGACTGCTTCCAGTGGGACTTTTCCACAACTGCAATGTATAATTTCTTC	757	
DB	1330	Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal	1344	
QY	758	AACAACTTC-----CATCATGATATTGAACGGTTTAAAGCCATT	796	
DB	1345	ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro	1364	
QY	797	AAGGATGCTTTGGTAGAATCCAGAGCTTTCAGCCAAAATTGGGAAGGCCCAATG	856	
DB	1365	GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro	1382	
QY	857	GGC-----AATGTCAGCCCCAGCACAACAAGGGTGCACACTGCAGGAGGTCAGGC	907	
DB	1383	GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly	1396	

QY	908	TCCTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAAAATATGTTGTTCTTCTATT	961	
DB	1397	MetPheAspSerSerValLeuCysGluCysTyrAspAla-----	1409	
QY	962	TCCAAATGCATGTTGTTGCAAAAATTATGAAGAAAGCCCAAGAAAGACACTAATGAGC	1021	
DB	1410	-----GlyCysAlaTrpTyrGluLeuThrPro-----AlaGluThrThrValArg	1424	
QY	1022	ATGCCAAACTACATGCAGACTGGAGTTG-----	1051	
DB	1425	LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp	1444	
QY	1052	GAAGC-----AGCATTACTCTGCCACCAACAGAAATTT	1084	
DB	1445	GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln	1464	
QY	1085	TCAGGA-----CTTCCAAGATTTCAGTC-----ACGATAGGCGGCTTCCTCA	1126	
DB	1465	SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG	1484	
QY	1127	TGCATCTCTCTGGGAGTGG-----TGGAGGCCACATGCCCTGCCTGCTGCTGCTCAG	1177	
DB	1484	nAlaProProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg	1499	
QY	1178	GGAGAAGAGGCCGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCTGGAG	1237	
DB	1499	gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu-----	1514	
QY	1238	GAATTTGGAGGTGCTTATCACAGATTCTCCACACTGAGT-----	1277	
DB	1515	-----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI	1531	
QY	1278	-----TTAAATCTAAGGGATTCAAAATGGAGTAGAGTATAAAG	1315	
DB	1531	eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG	1551	
QY	1316	TCTGAATGCA-----TGTTGATTTGCTTCTAGTCTAGAAATCTCTAGTTTAGAAAGAT	1369	
DB	1551	yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal-----	1569	
QY	1370	GTTTAGGGGAACATGAGGCTGCTGTCAGCAGCAACACAGGCTCCCTGCATCCCTGGGC	1429	
DB	1570	-----GlyArgValValLeuSerGlyLysProAlaIleProAspAr	1584	
QY	1430	CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCTTTTC	1483	
DB	1584	gLuValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI	1604	
QY	1484	TCAGGAGTGCATGCCCTGAGCATTGTTTGTCTGCGCCAGAGAGAGAGCTTGGGTCCC	1543	
DB	1604	eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu--	1621	
QY	1544	ATAGTCTGGGAGAGTGTCTGAGGCGCGGAGGCGCAGAGAGGCGCTTCGGAGAGCTC	1603	
DB	1622	-----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh	1638	
QY	1604	ACTCTGCTGCTCTT	1619	
DB	1638	rAsnTrpGlnLysLeu	1643	

RESULT 14
US-08-466-975A-23
; Sequence 23, Application US/08466975A
; Patent No. 5910404
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466.975A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/391.671
FILING DATE:
APPLICATION NUMBER: US 07/920.286
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-466-975A-23

Alignment Scores:
Pred. No.: 0.0196 Length: 2894
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-466-975A-23 (1-2894)
QY 641 TCTGTTTCCCC---TCGGATCACTTCCAGGACCACCAAAATACTTTCGCTGGG 697
DB 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
QY 698 TACTGTGACTGCTTTCGCCAGTGGGACATTGGCAACAACTCAATTGTAATAATTGTTC 757
DB 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
DB 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
QY 797 AAGGCATCTCTGGTAGAAATCCAGAACCTTCCAGCCCAAAANTTGGCAAGGCCAATTG 856
DB 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
QY 857 GGC-----AATGTCAAGCCAGCAGCAACAAAGGGTCAACTCGCAGGAGGTTCAGGC 907
DB 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510

QY 908 TGCCTGAAGAAT-----TACTGCGAGTGCTATAGAGCCCAAAATTATGTGTTCTTCTATT 961
DB 1511 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1523
QY 962 TGCAAAATGCATTGGTTGCAAAATATTAAGAAAGCCAGAACGACACACATATGAGC 1021
DB 1524 -----GlyCysAlaTyrPtyrGluLeuThrPro---AlaGluThrThrValArg 1538
QY 1022 ATGCCAAACTACATGCAGACTGGAGTTTG----- 1051
DB 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
QY 1052 GAAGGC-----AGCCATTACCTGCCACCACCAAGAAATTT 1084
DB 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGly 1578
QY 1085 TCAGGA-----CTTCCAAGATTCACTC-----ACGATAGCGCGCTTCCTCA 1126
DB 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG 1598
QY 1127 TGCATCTCCTGGGAGTGG-----TGAGGCCACATCGGCTGCTGCTGCTCAG 1177
DB 1598 nAlaProProSerProSerTyrAspGlnMetTrpLys-----CysLeuIleArg 1613
QY 1178 GCAGAAAGAGCGCGAGAAACACTGCTCCAAAGTCCCTGGCAGACAGATGATCCTGGAG 1237
DB 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
QY 1238 GAATTGGAGAGTCTTATACACAGATTCTCCACACTGAGT----- 1277
DB 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI 1645
QY 1278 -----TTAATCTAAGGATTGAAATCGAGTAGAGTATAAG 1315
DB 1645 eMetThrCysMetSerAlaAspLeuValValThrSerThrThrValLeuValGlyG 1665
QY 1316 TGTGAATGCA-----TGTGATTTTGTCTTAGTCTAGAAATCTCTAGTTAGAAAGGAT 1369
DB 1665 yValLeuAlaAlaLeuAlaTyrCysLeuSerThrGlyCysValValIleVal---- 1683
QY 1370 GTTTAGGGAACATGAGGCTGGCTGTCAGCAACACAGGCTCCCTGCTGCTGCTGGC 1429
DB 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
QY 1430 CCAGGAGTCTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC 1483
DB 1698 gLeuValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI 1718
QY 1484 TGAGGAGTGCATGCCCTGACGATTTGTTGCTGGCCCGAGAGAGAGAGCTTGGGTTCCC 1543
DB 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
QY 1544 ATAGCTGGGAGAGTGTCTCAGCGCGCGGCGGAGGAGAGAGAGGCTCGGAGAGCTC 1603
DB 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
QY 1604 ACTCTGTCGACTCTT 1619
DB 1752 rAsnTrpGlnLysLeu 1757
RESULT 15
US-08-391-671A-23
; Sequence 23, Application US/08391671A
; Patent No. 5922532
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON & VANDERHUYE P.C.
 STREET: 1100 NORTH GLEBE ROAD
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22201
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/391,671A
 FILING DATE: 21-FEB-1995
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 07/920,286
 FILING DATE: 14-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/EP91/02409
 FILING DATE: 13-DEC-1991
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: EP 90124241.2
 FILING DATE: 14-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: SADOFF, B.J.
 REGISTRATION NUMBER: 36,663
 REFERENCE/DOCKET NUMBER: 1487-5
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 7038164000
 TELEFAX: 7038164100
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2894 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-391-671A-23

Alignment Scores:			
Pred. No.:	.		
Score:	117.00	Length:	2894
Percent Similarity:	35.4%	Matches:	91
Best Local Similarity:	23.5%	Conservative:	136
Query Match:	3.06%	Mismatches:	438
DB:	2	Indels:	111
		Gaps:	23

US-09-743-237-3 (1-2134) x US-08-391-671A-23 (1-2894)

Qy	641	TCGCTTTCCCC---TCGGGATCAACTCTTTCCAGGACCCACCAAAATAACTTTGGCTGGG	637
Db	1424	SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly	1443
Qy	698	TACTGTGACTGCTTTGGCAGCTGGGACTTTTGAACAACACTGCAATTGTATAATTGTTGC	757
Db	1444	Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal	1458
Qy	758	AACAACWTG-----CATCATGATATTGAACGGTTTAAGCCATT	796
Db	1459	ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro	1478
Qy	797	AAGCAGTCGTGGTACAAAATCCAGAAGCTTTCAGCCAAAATAATGGGAGAGGCCCAATTG	856
Db	1479	GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro	1496
Qy	857	GGC-----AATGTCAAGCCCCAGCACACAAAGGGTGCAACTGCAGGAGGTCAAGC	907
Db	1497	GlyIleThrArgPheValAlaProGlyGlu-----ArgProSerGly	1510

Search completed: July 24, 2003, 14:39:13
Job time : 94.9246 secs

Qy	908	TGCTGAAGAAAT-----TACTGCGAGTGCTATGAGGCCCAAAATATGTGTTCTCTCTATT	961
Db	1511	MetPheAspSerValLeuCysGluCysTryAspAla-----	1523
Qy	962	TGCAAAATGCATTGTGTTCCAAAAATATGAGAAGAGCCCAACAGCAAGACACATAATGACG	1021
Db	1524	-----GlyCysAlaTrpTryGluLeuThrPro--AlaGluThrThrValArg	1538
Qy	1022	ATGCCAAATACATGCAGACTGGAGTTTG-----	1051
Db	1539	LeuArgAlaTrpMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp	1558
Qy	1052	GAAGGC-----AGCCATTACCTGGCCACCACCAAGAAATTT	1084
Db	1559	GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGly	1578
Qy	1085	TCAGGA-----CTTCCAAGATTACGTC-----ACGATAGCGCGCTTCTCTCA	1126
Db	1579	SerGlyGluAsnLeuProTryLeu-ValAlaTrpGlnAlaThrValCysAlaArgAlaG1	1598
Qy	1127	TGACTCTCCGCGAGGTGG-----TGGAGGCCACATGCGCGCTCGCTTGCTTCAC	1177
Db	1598	nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr	1613
Qy	1178	GGAGAAGAGCGCCGAGAAAGAACACTGCTCCAAGTGCTGCGCAGAGCAGATGATCCTGGAG	1237
Db	1613	gLeuLysProThrLeuHisGlyProThrProLeuLeuTryArgLeu-----	1628
Qy	1238	GAATTTGGAAAGGTCTTATCACAGATTCTCCACACTGAGT-----	1277
Db	1629	-----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTryI1	1645
Qy	1278	-----TTAATCTAAGGATTGAAATGGAGTAGAGATATAAG	1315
Db	1645	eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1	1665
Qy	1316	TGCAATGCA-----TGTGTATTCTTCTTAGTCTAGAAATCTAGTTTACAAAGGAT	1369
Db	1665	yValLeuAlaAlaLeuAlaAlaTrpCysLeuSerThrGlyCysValIleVal----	1683
Qy	1370	GTTTAGGGGAACATGAGCTGGCTGCTGCACACACACACAGGCTCCCTGCATCCCTGGGC	1429
Db	1684	-----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr	1698
Qy	1430	CCAGGGAGTTTACTACAGACTCTCT-----GAAGATGTGGCAACCACTCCCTTTTC	1483
Db	1698	gGluValLeuTryArgGluPheAspGluMetGluCysSerGlnHisLeuProTryI1	1718
Qy	1484	TGAGAGGTGCATGGCTGACGATTGTTTGTCTGGCCCGCAGGAGAGCTTGGTTC	1543
Db	1718	eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu--	1735
Qy	1544	ATAGTCTGGAGAGTCTCTCGAGGCGCGAGGGCGAGAGCGGCTCCGAGAGCTC	1603
Db	1736	-----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh	1752
Qy	1604	ACTCTGCTGCACTTT	1619
Db	1752	rAsnTrpGlnLysLeu	1757

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 24, 2003, 13:37:43 ; Search time 107.554 Seconds
(without alignments)
6298.674 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aatcggggtcaagcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/csqn2_1/USPTO_spool/US09743237/runat_24072003_113103_17652/app_query.fasta_1.4750
-DB=A_Geneseq_19Jun03 -QMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237.ecgn.1.1.376.e@runat_24072003_113103_17652 -NCPUP=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	43.1	299	21	AA668464
2	1251	32.7	295	21	AA668463
3	836.5	21.9	251	22	AA693348
4	730	19.1	147	22	AB86025
5	730	19.1	147	22	AA693330
6	670	17.5	528	22	ABG74678
7	483.5	12.6	950	22	ABG62035
8	441	11.5	438	20	AAW83392
9	378.5	9.9	280	22	ABG17958
10	278	7.3	53	22	AA676158
11	247.5	6.5	609	22	ABP70527
12	200	5.2	243	22	ABG68888
13	165	4.3	1057	24	ABP97370
14	158	4.1	1192	24	ABP97376
15	158	4.1	1207	24	ABP97377
16	158	4.1	1477	24	ABP97371
17	158	4.1	1535	24	ABP97374
18	158	4.1	1593	24	ABP97369
19	157.5	4.1	1251	24	ABP97375
20	157.5	4.1	1342	24	ABP97379
21	157.5	4.1	1512	24	ABP97372
22	157.5	4.1	1570	24	ABP97373
23	157.5	4.1	1628	24	ABP97368
24	149	3.9	4561	22	ABG30203
25	149	3.9	9222	22	ABG21064
26	130.5	3.4	2743	23	AB81598
27	130.5	3.4	3695	23	AB81588
28	130.5	3.4	3596	23	AAE17310
29	130.5	3.4	3705	23	AAE17309
30	130	3.4	11300	23	AB84277
31	129.5	3.4	1403	24	ABJ37903
32	129.5	3.4	19338	24	ABP76681
33	128	3.3	19338	24	AB898398
34	126	3.3	1398	24	ABJ37900
35	126	3.3	1404	24	ABJ37899
36	126	3.3	1577	24	ABJ37904
37	125	3.2	1037	22	AAU07142
38	124	3.2	3011	13	AA21519
39	122.5	3.2	19938	24	ABP76679
40	122.5	3.2	19938	24	ABP76679
41	121	3.2	507	16	AA66631
42	121	3.2	1017	22	AA659813
43	121	3.2	1615	22	AA659826
44	120	3.1	3010	14	AA30616
45	120	3.1	3011	14	AA31621

ALIGNMENTS

RESULT 1
AA668464

ID AA668464 standard; Protein; 299 AA.

XX AA668464;

AC AA668464;

XX 25-APR-2000 (first entry)

XX Human testis specific factor tesmin SEQ ID NO:5.

DE Testis specific factor; tesmin; cell death; regulation; spermatocyte; differentiation regulatory factor; male germ cell regulatory actor;

XX germ cell differentiation; sterility.

OS Homo sapiens.

XX WO200004147-A1.

XX

PD 27-JAN-2000.
 XX
 PF 16-JUL-1999; 99WO-JP03859.
 XX
 PR 17-JUL-1998; 98JP-0219856.
 XX
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 PI
 PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX
 DR WPI: 2000-147785/13.
 DR N-PSDB; AAZ88156.
 XX
 PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 XX
 PS Claim 1; Page 50-52; 63pp; Japanese.
 XX
 CC The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 299 AA;
 Alignment Scores:
 Pred. No.: 7.44e-151 Length: 299
 Score: 1647.00 Matches: 299
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.07% Indels: 0
 DB: 21 Gaps: 0

US-09-743-237-3 (1-2134) x AAY68464 (1-299)

Qy 407 ATGTGTATGCCAATGAAAGGGGACACAAATGCTATATAGACAATCTAGACA 466
 Db 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
 Qy 467 AGAGAACTAAAGCACCTCATTGGTTCTCCTCAGATCAAGATCAAAATAATCTACAG 526
 Db 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGln 40
 Qy 527 TCAGATGTCCCTAAACCAATGACTCTTTAGTAGGGAGATTTTCCAGCATCAACAAAA 586
 Db 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
 Qy 587 TTAATCTCATACACAACACTTTCAGGAGGACCTTACCATCGGTAGTCAAGGGTCTGCT 646
 Db 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
 Qy 647 TTCCTCCGGATCAACTCTCCAGGACCCACAAATAACTTTGGCTGGTACTGTGAC 706
 Db 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
 Qy 707 TGCCTTTCAGTGGGAGCTTTTGCACAACTGCAATTTGTAATAATTTGTGCAACAACCTG 766
 Db 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnCysAsnCysAsnLeu 120
 Qy 767 CATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGCTTGTGTAGAAATCCAGAGCT 826
 Db 121 HishisAspIleGluArgPheLysAlaIleLysAlaCysLeuGlyArgAsnProGluAla 140
 Qy 827 TTCAGCCAAAATTTGGGAAGGGCAATTTGGGCAATGTCAAGCCCAACCAAGGG 886
 Db 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
 Qy 887 TGCAACTGCAGGAGGTGAGGCTGCTGAAGAATTTACTGCGAGTGTCTATGAGGCCCAAT 946
 Db 161 CysAsnCysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGlnIle 180
 Qy 947 ATGTGTTCTTCTATTTCGAATGTCATTTGGTTGCAAAATTTATGAGAAAGCCCAAGCA 1006

Db 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluSerProGluArg 200
 Qy 1007 AAGACACTAATGAGCATGCCAACTACATGCAGACTGGAGCTTTGGAGGAGCCATTAC 1066
 Db 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
 Qy 1067 CTGCCACACAGAAATTTTCAGGACTTCCAAAGATTACGATCAGTAGCGGCTTCCTCA 1126
 Db 221 LeuProThrLysPheSerGlyLeuProArgPheSerHisAspArgProSerSer 240
 Qy 1127 TGCATCTCTGGAGGTGGTGGAGCCACATGCCCTGCCCTGCTGCTCAGGAGAGAG 1186
 Db 241 CysIleSerTrpGluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGlu 260
 Qy 1187 GCCGAGAAAGAACACTGCTCCAAGTGGCTGCCAGACAGATGATCCTCGAGGAATTTGGA 1246
 Db 261 AlaGluLysGluHisCysSerLysCysLeuAlaGlnMetIleLeuGluGluPheGly 280
 Qy 1247 AGGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGATTGAAATGGAG 1303
 Db 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 2
 AAY68463
 ID AAY68463 standard; Protein: 295 AA.
 XX
 AC AAY68463;
 XX
 DT 25-APR-2000 (first entry)
 XX
 DE Mouse testis specific factor tesmin SEQ ID NO:4.
 XX
 KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility.
 XX
 OS Mus musculus.
 XX
 PN WO200004147-A1.
 XX
 PD 27-JAN-2000.
 PF 16-JUL-1999; 99WO-JP03859.
 XX
 PR 17-JUL-1998; 98JP-0219856.
 XX
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX
 PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX
 DR WPI: 2000-147785/13
 DR N-PSDB; AAZ88155, AAZ88157.
 XX
 PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 XX
 PS Claim 1; Page 47-49; 63pp; Japanese.
 XX
 CC The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 295 AA;
 Alignment Scores:
 Pred. No.: 2.23e-112 Length: 295
 Score: 1251.00 Matches: 228
 Percent Similarity: 84.95% Conservative: 26
 Best Local Similarity: 76.25% Mismatches: 41
 Query Match: 32.71% Indels: 4

```
DB:                21                Gaps:                1
US-09-743-237-3 (1-2134) x AAY68463 (1-295)
QY 407 ATGGTGATATGCCAAATGGAAGGGGCACAAATGCTATGTATAGACAATTCAGACA 466
Db 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 467 AGAGAACTAAAGCACTCCATTTGGTTCCCTCAGTATCAAGATCAAAATAATATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerSerPheProGln 40
QY 527 TCAGATGTCCTTAAACAATGACTGCTTGTAGTGGAGATTTTGGCAGCATCAACAAA 586
Db 41 SerGluLeuProLysPheMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 587 TTAATCTCATTAACACAACACTTGAGGAGCCTTACCATCGGTAGTCAACGGTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 647 TTCCTCTCGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGAC 706
Db 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 707 TGCTTGGCCAGTGGGACTTTTGCACAACACTGCAATTTGTAATTTGTCACAACACTTG 766
Db 101 CysPheSerSerGlyAspPheCysAsnSerCysSer-----CysAsnAsnLeu 116
QY 767 CATCATGATATTAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAATTCAGAAGCT 826
Db 117 ArgHisGluLeuGluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAla 136
QY 827 TTCAGCAAAATAATGGGAAGGCGCAATTTGGCAAGTCTCAAGCCCGCACCAACAAGGG 886
Db 137 PheGlnProLysMetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGly 156
QY 887 TGCAGTGCAGAGGTGAGCTGCTGCAAGAAATTTACTGCGAGTGTATGAGGCCCAATT 946
Db 157 CysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIle 176
QY 947 ATGTGTTCTTCTATTGCAAAATTCATTGTTGCAAAATTTATGAAGAAAGCCAGAACGA 1006
Db 177 MetCysSerSerIleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArg 196
QY 1007 AAGACACTATGAGCATGCCAACTACATGCAGACTGAGGTTTGGGAAGCAGCCATTAC 1066
Db 197 LysMetLeuMetSerThrProHisTyrMetGluProGlyAspPheGluSerSerHisTyr 216
QY 1067 CTGCCACCAACGAAATTTTCAGGACTTCCAAGATTCAGTCACGATAGGCGGCTTCCTCA 1126
Db 217 LeuSerProAlaLysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSer 236
QY 1127 TGCATCTCCGGAGGTGGAGGCCACATGCGCTGCTGCTGCTCAGGGAGAGAG 1186
Db 237 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 256
QY 1187 GCCGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCTGGAGGAATTTGGA 1246
Db 257 AlaGluGlnGluHisCysSerProSerLeuAlaGluGlnMetIleLeuGluPheGly 276
QY 1247 AGGTGCTTATCAGAGATTCACACTGAGTTTAAATCTAAGGGATTGAATGGAG 1303
Db 277 ArgCysLeuSerGlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295
RESULT 3
AAM93348
ID AAM93348 standard; Protein: 251 AA.
XX
XX AC AAM93348;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2895.

XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
OS Homo sapiens.
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
DR N-PSDB: AAK94268.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PT
PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 251 AA;

Alignment Scores:
Pred. No.: 3,93e-72 Length: 251
Score: 836.50 Matches: 167
Percent Similarity: 95.43% Conservative: 0
Best Local Similarity: 95.43% Mismatches: 5
Query Match: 21.88% Indels: 3
DB: Gaps: 1

US-09-743-237-3 (1-2134) x AAM93348 (1-251)
QY 9 GTCAGCGGAGCTCGCGGGGGGCACAGCGCGGGGAGCTCTCGGGAGTACCCC 68
Db 79 ValIysAlaLysLeuAlaGlyGlyAspSerAspGlyGlyGluLeuLeuGlyTyrPro 98
QY 69 GGGATCCAGAGCTCAGCGCGCTGAGAGCTCGCGTCTCTGAGGCGCCCGCAGCGGCC 128
Db 99 GlyIleProGluLeuSerAlaLeuGluAspValAlaLeuLeuGluAlaProGluPro 118
QY 129 GCCTGCAACGTGCACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 188
Db 119 AlacysAsnValHisPheLeuSerSerLeuLeuProAlaHisArgSerProAlaValLeu 138
QY 189 TGCCCTTGGGCGCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 248
Db 139 ---ProLeuGlyAlaIleValLeu-GluGlyAlaSerHisProGlyValArgMetIlePr 157
QY 249 AGTTGAATCAAGTAAGAGGTGCTACTACTACAAGTAATAATCCGGAAGCAACTT 308
Db 157 oValGluIleLysGlu-AlaGlyGlyThrThrSerAsnAsnProGluGluAlaThrL 177
```


PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 03-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483232/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 PS Claim 11; SEQ ID NO 1409; 766pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a protein of the
 CC invention.
 XX
 SQ Sequence 147 AA;
 Alignment Scores:
 Pred. No.: 6.95e-62 Length: 147
 Score: 730.00 Matches: 144
 Percent Similarity: 97.30% Conservative: 0
 Best Local Similarity: 97.30% Mismatches: 3
 Query Match: 19.09% Indels: 1
 DB: 22 Gaps: 0
 US-09-743-237-3 (1-2134) x ABB96025 (1-147)
 QY 229 CCGGCGCTCCGATGATCCAGTGAATCAAGTAAGCAGGTGGTACTACTACAAGTAA 288
 DB 1 ProglyValArgMetIleProValGluIleLysGlu-AlaGlyGlyThrThrSerAs 20
 QY 289 TAATCCGGAAGAAGCAACTTTGTCAGAAATCTTCTGCTCAGGAATCTCTGTCAGAGTTCCC 348
 DB 20 nAsnProGluGluAlaThrLeuGlnAsnLeuLeuAlaGlnGluSerCysCysLysPhePr 40
 QY 349 ATGGTCCCGAAGTACTAGAGATGCCCTGCTGCTTCTTAAGAAAGATTCCAAACCAAT 408
 DB 40 oSerSerGlnGluLeuGluAspAlaSerCysCysSerLeuLysLysAspSerAsnProMe 60

QY 409 GGTGATATGCAATTGAAGGGGCACACAAATGCTATGATAGACAATCTAGACAAG 468
 DB tValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThrAr 80
 QY 469 AGAATAAAAGCACTCCATTTGGTTCCTCAGTATCAAGATCAAAATTAATTATCTACAGTC 528
 DB gGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnse 100
 QY 529 AGATGTCCTTAAACCAATGACTGCTTTAGTAGGAGAGATTTTGGCCAGCATCAACAAATTT 588
 DB rAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLysLe 120
 QY 589 AAATCTCATTACACAACTTGAGGGAGCGCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
 DB uAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAlaPh 140
 QY 649 CCCCTCGGGATCAACTCTTCCA 670
 DB 140 eProSerGlySerThr***Pro 147
 RESULT 5
 AAM95330
 ID AAM95330 standard; Protein; 147 AA.
 AC AAM95330;
 XX
 XX 21-NOV-2001 (first entry)
 DT
 XX Human reproductive system related antigen SEQ ID NO: 3988.
 DE Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy.
 KW
 XX Homo sapiens.
 OS
 XX WO200155320-A2.
 PN
 XX 02-AUG-2001.
 PD
 XX 17-JAN-2001; 2001WO-US01339.
 PF
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.

Qy 409 GGTGATATGCCAATTGAAGGGGCGCACAAATGCTATGATATAGACAAATCTAGACAAAG 468
 Db 60 TValleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThrAr 80
 Qy 469 AGAACTAAAGCACTCCATTGTTCTCAGTCAAGATCAAGATCAAAATATATCTACAGTC 528
 Db 80 gLlueuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSe 100
 Qy 529 AGATGTCCTCCAAACCAATGCTGCTTTAGTAGGAGAGATTTTGGCAGCATCAACAAAT 588
 Db 100 rAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLysLe 120
 Qy 589 AAATCTCATTAACAACAACCTGAGGAGCGCTTACCATCGGTAGTCAACGGGTCTGTTT 648
 Db 120 uAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAlaPh 140
 Qy 649 CCCCTCGGGATCAACTCTTCCA 670
 Db 140 eProSerGlySerThr***Pro 147
 RESULT 6
 ID ABG74678 standard; Protein: 528 AA.
 XX AC ABG74678;
 XX DT 10-MAY-2003 (first entry)
 XX DE Human CGDD protein 2369279CD1 SEQ ID 4.
 XX KW Human; cell growth, differentiation and death; CGDD; cardiant; cancer;
 KW cytostatic; neuroprotective; gene therapy; diagnosis; cardiovascular;
 KW neurological disorder.
 XX OS Homo sapiens.
 XX PN WO2003014322-A2.
 XX PD 20-FEB-2003.
 XX PE 08-AUG-2002; 2002WO-US25465.
 XX PR 08-AUG-2001; 2001US-311017P.
 PR 17-AUG-2001; 2001US-313070P.
 PR 17-AUG-2001; 2001US-313071P.
 PR 24-AUG-2001; 2001US-314678P.
 PR 31-AUG-2001; 2001US-316692P.
 PR 07-SEP-2001; 2001US-317913P.
 PR 14-SEP-2001; 2001US-322182P.
 PR 07-DEC-2001; 2001US-340747P.
 PR 20-DEC-2001; 2001US-342761P.
 PR 29-MAR-2002; 2002US-369129P.
 XX PA (INCY-) INCYTE GENOMICS INC.
 XX PI Azimzal Y, Barroso I, Baughn MR, Becha SD, Borowsky ML, Duggan BM;
 PI Elliott VS, Emerling BM, Forsythe IJ, Gietzen KJ, Gorvad AE;
 PI Graul RC, Griffin JA, Gururajan R, Hafalia AJA, Ison CH, Kable AE;
 PI Khan FA, Lee S, Lee SY, Li JX, Reddy R, Richardson TW;
 PI Sprague WW, Swarnakar A, Tang YT, Warren BA, Xu Y, Yao MG, Yue H;
 Yue H;
 XX WPI; 2003-268197/26.
 DR N-PSDB; ABQ77417.
 XX PT New polypeptide associated with cell growth, differentiation and death,
 PT useful for preparing a composition for diagnosing or treating a disease
 PT e.g., cardiovascular or neurological disorders or cancer
 XX PS Claim 59; Page 201-202; 267pp; English.
 XX CC This invention describes a novel polypeptide associated with cell growth,
 CC differentiation and death (CGDD) which has cardiant, cytostatic and

CC neuroprotective activity. The polypeptides of the invention are useful
 CC for preparing a composition for diagnosing, or treating a disease or
 CC condition associated with decreased expression or overexpression of
 CC functional CGDD e.g., cardiovascular or neurological disorders or cancer.
 CC The polynucleotides of the invention can be used for gene therapy.
 CC ABG74675-ABG74701 represent the human CGDD polypeptides encoded by
 CC ABQ77414-ABQ77440.
 XX
 SQ Sequence 528 AA;
 Alignment Scores:
 Pred. No.: 8.19e-56 Length: 528
 Score: 670.00 Matches: 171
 Percent Similarity: 49.23% Conservative: 52
 Best Local Similarity: 37.75% Mismatches: 128
 Query Match: 17.52% Indels: 102
 DB: 24 Gaps: 16
 US-09-743-237-3 (1-2134) x ABG74678 (1-528)
 Qy 113 GGCCCGCGAGCGCCGCTGCAAGTGCACCTTCTGCTCTGCTACCGCGCACCG 172
 Db 109 GlyValSerThrSerGlnPheLysThrIleIleProLeuAlaThrAlaProAsnValGln 128
 Qy 173 CAG-----CCCGCGGGTGTGTTTGGCCCTGGGGCGGCTGCTCGAAGAGAGCCT 223
 Db 129 GlnIleGlnValProGlySerLysPheHisTyrValArgLeuValThrAla----- 145
 Qy 224 CCCACCGGGGCTCGCATGATCCAGTGTGAATCAAGTAAAGAGGTGGTACTACTACA 283
 Db 146 ---ThrSerAlaSerSerSerThrGln-----ProVal 155
 Qy 284 AGTAATAATCCGGAAGAAGCAACTTGCAGACATCTCTTGTCTCAGGAATCCTGTTGCAAG 343
 Db 156 SerGlnAsnProSerThrAsnThr----- 163
 Qy 344 TTCCCATGGTCCCAGGAACCTAGAGATGCTCTGCTCTTCTCTTAAG----- 391
 Db 164 -----GlnProLeuGlnAlaLysProValValValValValValValValValVal 179
 Qy 392 AAAGATTCCAAACCAATGGTGTATGCGAA-----TTG 424
 Db 180 ArgMetSerValProIleValSerAlaGlnAlaValLysGlnValValProLysProIle 199
 Qy 425 AAAGGGGCGCACAAATGCTATGTATAGACAATCTTACACAAGAGAACTAAAGCACATC 484
 Db 200 AsnProThrSerGlnIleValThrSerGlnProGlnGlnArgLeuIleMetProAla 219
 Qy 485 CATTGTGTTCTCAGTATCAAGATCAAAATAATATCTACAGTCAGAT-----GTC 535
 Db 220 ThrProLeuProGlnIleGlnProAsnLeuThrAsnLeuProProGlyThrValLeuAla 239
 Qy 536 CCTAAACCAATGACTGCTTTAGTAGGG---AGATTTTGGCCAGCATCAACAAATTAAT 592
 Db 240 ProAlaProGlyThrGlyAsnValGlyTyrAlaValLeuProAla-----Gln 255
 Qy 593 CTCATTACACAAACACTTGGAGGAGCCCTACCTAGTAGTCAACGGGTCTGCTTTC--- 649
 Db 256 TyrValThrGlnLeuGlnGlnSerSerTyrValSerIleAlaSerAsnSerThrPheThr 275
 Qy 650 -----CCCTCGGGA 658
 Db 276 GlyThrSerGlyIleGlnThrGlnAlaArgLeuProPheAsnGlyIleIleProSerGlu 295
 Qy TCA---ACTCTTCCAGGACCACTA-----AAATAACTTTTGGTGGGTAC 700
 Db 296 SerAlaSerArgProArgLysProCysAsnCysThrLysSerLeuCysLeuLysLeuTyr 315
 Qy 701 TGTGACTGCTTTGGCAGTGGGACTTTTGCACAACTGCAATTTAAATAATTTGTTGCAAC 760
 Db 316 CysAspCysPheAlaAsnGlyGluPheCysAsnAsnCysAsnCysThrAsnCysTyrAsn 335
 Qy 761 AACTTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGGCATGTCTTGGTAGAATCCA 820


```
Db 576 uilleAsnArgThrAlaAsnValGlnArgValSerSerSerSerSerProSerSerAs 596
Qy 382 TTCTCTTAAAGAGATTCCCAACCAATGGTGTATGCAATTAAGAGGGGCACACAAAT 441
Db 596 nSerThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 616
Qy 442 GCTATGT-----ATGACAAATCTTAGAACAAGAACTAAAGCAGCTCCATTGGT 492
Db 616 tileCysAsnSerGlyValProGlnSerSerThrSerAlaMetArg-----631
Qy 493 TCCTCAGTATCAAGATCAAAATAATATCTACAGTCAGATGTC-----CCTAAACCAAT 546
Db 632 -----ArgGlyThrThrGlyThrValGluAsnLysThrArgArgProProIle 648
Qy 547 GACTGCTTTAGTGGGAGATT-----568
Db 648 eSerProGlnGlnHisArgPheLysGlnMetGlyProGlnGlnSerLysHisGlnG 668
Qy 569 -TTGCCAGCATCAACAAATAATCTCATTTACACAACTT-----610
Db 668 nLeuGlnAlaGlnAlaLysGlnArgIleArgGlnGlnLeuProThrGlnSerTh 688
Qy 611 -----GAGGAGCCTTACCATCGTA-----GTCACCGGTC 642
Db 688 rProIleLysValGluProLysLeuProThrLeuProProGlyValLysAlaAsnValPr 708
Qy 643 TGCTTTCCCTCGGATCAACTCTTCAGGACCAACCAATAACTTTGGCTGG-----697
Db 708 oAlaLysProLeuPheGluValLeuLysProProAlaThrAlaAlaAlaGlyAlaVa 728
Qy 697 -----697
Db 728 lAspProLeuGlyGlyMetThrSerArgArgLysHisCysAsnCysSerLysSerGlnCy 748
Qy 698 -----TACTGTACATGCTTGGCTGAGGACTTTTGCACAACTGCAATTGTAA 747
Db 748 sLeuLysLeuTyrcysAspCysPheAlaAsnGlyGluPheCysGlnAspCysThrCysLy 768
Qy 748 TAATTGTGCAACAACTTGATCATATGACGGTTTAAAGCCATTAAGGCATGCT 807
Db 768 sAspCysPheAsnAsnLeuAspTyrgluValGluArgGluArgAlaIleArgSerCysLe 788
Qy 808 TGTAGAAATCCAGAGCTTTCAGCCCAAAATTTGGAAGGCCAATTTGGCAATGCCA 867
Db 788 uAspArgAsnProSerAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetAr 808
Qy 868 GCCCCAGCACAAAGGGTCAACTGCAGAGGTGCTGAGGAGTCTGAGAGTACTCGCA 927
Db 808 g----LeuHisAsnLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrcysGl 827
Qy 928 GTGCTATGAGGCCCAATATGCTTCTTCTTATTTGCAATGCAATGCTGTTGCAAAATTA 987
Db 827 uCysTyrgluAlaLysIleProCysSerSerIleCysLysCysValGlyCysArgAsnMe 847
Qy 988 TGAAGAAGCCAGCAAGAAAGACACTAATGACATGCCAACTACATGACAGTGGAGG 1047
Db 847 tGluAspArgProAsp-----ValaspMetAspSerLeuAspGlyLeuMetGl 863
Qy 1048 TTTGGAAGCGAGCCATTACCTGCCACCAACAAATTTTTCAGGACTTCCAGATTTCAGTCA 1107
Db 863 yValGluGlyGln-----LysLysAspLysAlaLysAsnLysGlnLeuAs 878
Qy 1108 CGATAGCGGCTTCTTCATCATCTCTGCGGAGGTGGTGGAGGCCACATGCGCGCTGCCT 1167
Db 878 nGluAsnArgAlaAsnIleTyrgPheThrAspAspValIleGluAlaThrIleMetCysMe 898
Qy 1168 CTTTCTCAGGAGAGAGAGCCGAGAAAGAACACTGCTCCAAAGTGCCTGCGCAGAGAGAT 1227
Db 898 tileSerArgIleValIleMetHisGluLysGlnAsnValAlaValGluAspMetGluArgGl 918
Qy 1228 GATCTCGGAGAAATTTGGAAGTGCCTTATCACAGATTCTC 1267
Db 918 uValMetGluGluMetGlyGluSerLeuThrThrGlnIleIle 931
```

```
RESULT 8
AAW83392
ID AAW83392 standard; Protein; 438 AA.
XX
AC AAW83392;
XX
DT 29-MAR-1999 (first entry)
XX
DE Caenorhabditis elegans synMuv protein LIN-54.
XX
KW LIN-54; synthetic multivulvar; synMuv; signal transduction;
KW animal model; tumour suppressor; retinoblastoma; cancer;
KW cell proliferation; gene therapy.
XX
OS Caenorhabditis elegans.
XX
PN W09854299-A1.
XX
PD 03-DEC-1998.
XX
PF 28-MAY-1998; 98WO-US11043.
XX
PR 28-MAY-1997; 97US-0047996.
XX
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Ceol C, Horvitz HR, Lu X;
XX
DR WPI; 1999-045362/04.
XX
DR N-PSDB; AAV72865.
XX
Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
useful for treating diseases associated with altered levels of cell
proliferation, e.g. carcinomas
XX
PS Claim 7; Fig 13; 70pp; English.
XX
This is the amino acid sequence of LIN-54, a novel protein of
Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
synthetic multivulvar (synMuv) gene involved in cell fate and cell
proliferation, and is part of a pathway that may be used as a
genetic and biochemical model system for tumour suppression and
cancer in mammals. synMuv pathway genes and proteins may be used
to identify genes which are part of the mammalian pathway and
to identify genes, proteins and therapeutic compounds which
modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
C. elegans synMuv polypeptides selected from LIN-37, -35, -55, -52,
-53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
(1) nucleic acids (see AAV72866-67) encoding the mouse and human
homologues of C. elegans LIN-54; (2) vectors containing the nucleic
acids; (3) transgenic cells; (4) a pure mammalian synMuv
polypeptide, and (5) an antibody which binds to a synMuv family
protein. The synMuv nucleic acids and polypeptides can be used to
diagnose and treat, especially by gene therapy, conditions
involving altered levels of cell proliferation, e.g. synMuv-
associated carcinomas.
XX
SQ Sequence 438 AA;
```

```
Alignment Scores:
Pred. No.: 1,35e-33 Length: 438
Score: 441.00 Matches: 93
Percent Similarity: 55.31% Conservative: 32
Best Local Similarity: 41.15% Mismatches: 61
Query Match: 11.53% Indels: 40
DB: 20 Gaps: 9
```

US-09-743-237-3 (1-2134) x AAW83392 (1-438)

```
Qy 698 TACTGTGACTGCTTGGCAGTGGGAGCTTTTGGCAACAACTGCAATTGTAATTAATTGTTC 757
Db 191 TycCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHis 210
```


Db 232 GluSerValGluSerArgThrSerGly- MetThrIleGlnArg 245
AAG76158
Qy 1160 GCCTGCTGCTGCTCAGGAGAGAGCCGAGAGAACACACTGCTCCAAGTGCCTGGCA 1219
Db 246 SerCysLeuAlaLeuGlnHisThrAspLeuGlu- CysSerIleTyrGly 262
Qy 1220 GAGCAGATGATCCTCGGAGGAATTTGGAAGGTGCTTATCAGAGATTCCTCCACACTGAGCTTT 1279
Db 263 LeuGlnLysAlaLeu-ArgAsnLeuHis- TyrGlnArgThrArgThrAlaGlnLe 280
Qy 1280 A 1280
Db 280 u 280
RESULT 10
ID AAG76158 standard; Protein: 53 AA.
XX AAG76158;
XX 03-SEP-2001 (first entry)
XX Human colon cancer antigen protein SEQ ID NO:5922.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 11.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX N-PSDB; AAH35563.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 8365; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAH7789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX Sequence 53 AA;

Alignment Scores:
Pred. No.: 3.71e-18 Length: 53
Score: 278.00 Matches: 53
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.27% Indels: 0
DB: 22 Gaps: 0
US-09-743-237-3 (1-2134) x AAG76158 (1-53)
Qy 1145 GTGAGGCCACATCGCGCTGCTGCTCAGGAGAGAGCCGAGAGAACACACTGC 1204
Db 1 ValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGluAlaGluLysGluHisCys 20
Qy 1205 TCCAAAGTGCCTGGCAGAGCAGATGATCCTGAGGAAATTTGGAAGGTGCTTATCAGAGATT 1264
Db 21 SerLysCysLeuAlaGluGlnMetIleLeuGluPheGlyArgCysLeuSerGlnIle 40
Qy 1265 CTCACACTGAGTTTAAATCTAAGGGATTGAAAATGGAG 1303
Db 41 LeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 53
RESULT 11
ABP70527
ID ABP70527 standard; Protein: 609 AA.
XX AC ABP70527;
XX DT 22-APR-2003 (first entry)
XX Histone deacetylase chromatin-associated protein complex subunit.
XX DE Histone acetyltransferase; histone deacetylase; gene expression profile;
XX KW Chromatin-associated protein; gene expression.
XX OS Arabidopsis thaliana.
XX PN WO2003000715-A1.
XX PD 03-JAN-2003.
XX PF 21-JUN-2002; 2002WO-US19750.
XX PR 22-JUN-2001; 2001US-300135P.
XX PA (CERE-) CERES INC.
XX PI Dang V, Okamuro J;
XX DR WPI; 2003-175280/17.
XX PT New chimeric polypeptide comprising a histone acetyltransferase
XX polypeptide segment and a segment comprising a histone deacetylase
XX chromatin-associated protein complex subunit, useful for modulating
XX gene expression in cells -
XX PS Disclosure; Page 61; 85pp; English.
XX CC The specification describes chimeric histone acetyltransferase
XX polypeptides. The chimeric polypeptides comprise a polypeptide segment
XX that exhibits histone acetyltransferase activity, and a polypeptide
XX segment having 40% or greater sequence identity to a subunit of a
XX histone deacetylase chromatin-associated protein complex. The chimeric
XX polypeptides are useful for determining gene expression profiles in
XX specific cells, for modulating gene expression in specific cells, and
XX for making genetically modified eukaryotes. The present sequence
XX represents an exemplary histone deacetylase chromatin-associated protein
XX complex subunit, which can be used to produce chimeric polypeptides of
XX the invention.
XX SQ Sequence 609 AA;
Alignment Scores:

Db 349 -----TrpCysArgCys 352
 QY 931 CTATGAGGCCAAATATATGTG----- 952
 Db 352 sGlnAlaGlyGlnValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlu 372
 QY 953 -----TCTTCTATTTCG---AAATCATTTGGTTGCAAAATATAT---GAAGA 993
 Db 372 nValThrGluArgGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGlu 392
 QY 994 AGCCCAAGACGAAACACACTAATGACATGCCAAACTACATCCAGACTGGAGTTTGGGA 1053
 Db 392 uHisPro-----GL 395
 QY 1054 AGCAGCCATTACCTGCCACCAACGAAATTTTCAGGACTTCCAAGATTTCAGTCACCATAG 1113
 Db 395 uGlySerArgTrpValPro-----Asp 404
 QY 1114 GCGGCTCTCTCATGATCTCTCTGGAGGTGTGGAGCCACATCGCC-----TG 1164
 Db 404 rAlaCysSerSerCysValCysHisGlyGlyVal---ValThrCysAlaArgIleGlnCys 423
 QY 1165 CCTG-----CTTGCTCAGGAGACAGCGCCGAGAAAGAACACTGCTCCAAAGTCCCTGCC 1218
 Db 423 sIleSerSerCysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 441
 QY 1219 AGAGCAGATGATCTCTGGAGAAATTTGGAAGTCTTATCACAGATTCTCCACACTCAGTT 1278
 Db 442 -----SerAspCys 444
 QY 1279 TAATCTAAGGANTGAAATGAG-----TAGAGTATAAAGTG 1317
 Db 444 sGluHisGluGlyArgGlySerGlyGluProGlyGluSerPheGlnProGlyAlaAspProCys 464
 QY 1318 TGAATCATGCTGATTTTGTCTAGTCTAGAAATCTAGTTAGAAAGATGTTTAGG 1377
 Db 464 sGluValCysIle----- 468
 QY 1378 GAACATGAGCTGGCTCTGCAGCAACACAGAGCTCCCTGCATCCCTGGGCCAGGGAG 1437
 Db 469 -----CysGluProGlnProGlyGlyProProSerLeu----- 479
 QY 1438 TTTACTCAGAGCTCTCTGAGATGT---GGCAACCATGCCCTTTCTGAGGA----- 1489
 Db 480 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysPr 493
 QY 1490 -----GGTGCATGGCTGAGCATTTCT----- 1510
 Db 493 oProSerGlnLeuLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLe 513
 QY 1511 -----TTGCTGGCCCA----- 1522
 Db 513 uSerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysThrTh 533
 QY 1523 -----GAGGAGAGCTTTGGTTTCCATA----- 1546
 Db 533 rCysGlnCysGlnAspLeuThrTriLeuCysIleHisGlnAlaCysProGluLeuSerCys 553
 QY 1547 -----GTCTGGGAGAGTGCT 1563
 Db 553 sProLeuSerGluArgHisThrProProGlySerCysCysProValCysArgGluCysVa 573
 QY 1564 GCAGGCGCGGAGGAGCAGAGCGCCCTCGGAGAGCTCACTGCTGGTCTTCTTCCTC 1623
 Db 573 lVal-GluAlaGluGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProS 591
 QY 1624 TCAGAAATGT-----TGCTCTGGAGGC-----T 1647
 Db 591 eAsnAlaCysIleAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluGluC 611
 QY 1648 GCTCTGCTGATAAACCCTTAATGTTTGTGTTTTCATTAATTTAGAAATAAGTT 1707
 Db 611 ysGlnAla-----LeuSerCysProHisGlyTrpAlaLysValProG 625

QY 1708 CTCGGATGGGTGTGTGATACCACTTAAATCTCTAGAGAACTACTAGAACCTAAAG 1767
 Db 625 luAlaasp-SerCysCysGluArg----- 632
 QY 1768 ATTTCTGTAGCGTAGATATTTCCCA-----GAGACACGCAACTGTCTCAGTCTTT 1818
 Db 633 -----CysGlnAlaProThrGlnSerCysValHisGlnGlyArgGluVal----- 647
 QY 1819 CCTAAGGCCCCGGGAGAGCGCAGGCAATGGGGCTCGCAGGCCAGCTTGCACCAACATG 1878
 Db 648 -----AlaSerGlyGluArg-----TrpThrVal-----AspThrCysThrSerCys 661
 QY 1879 TCTTGATGTAGAGGACTTAAATATTCAGTTTCTTCTGTGTCTTACTTGAATTTGGA 1938
 Db 662 SerCysMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 681
 RESULT 14
 ID ABP97376
 XX ABP97376 standard; Protein; 1192 AA.
 AC ABP97376;
 XX 10-MAY-2003 (first entry)
 DT
 XX Human kielin-like protein NHP9, SEQ ID NO:18.
 DE
 XX Human: kielin-like protein; NHP: chromosome 7; development;
 KW signal transduction; kielin-related disorder; diagnosis; treatment;
 KW drug screening; clinical trial monitoring; cosmetic; nutraceutical.
 OS Homo sapiens.
 PN WO2003004609-A2.
 XX 16-JAN-2003.
 XX 03-JUL-2002; 2002WO-US21118.
 XX 03-JUL-2001; 2001US-302949P.
 PR 29-AUG-2001; 2001US-315634P.
 XX (LEXI-) LEXICON GENETICS INC.
 XX Walke DW, Scoville J, Turner CA;
 XX WPI: 2003-210352/20.
 DR N-PSDB; ABZ75047.
 XX New human kielin-like nucleic acids, useful for preparing a composition
 PT for treating human kielin-related disorders -
 XX Claim 3; Page 71-74; 85pp; English.
 CC The invention relates to novel human kielin-like proteins (referred to as
 CC NHPs; ABP97368-ABP97379) and cDNAs encoding them (ABZ75039-ABZ75050). The
 CC invention also encompasses recombinant expression vectors and host cells
 CC comprising a nucleic acid of the invention. The NHP cDNAs were obtained
 CC by aligning cDNAs from human kidney, foetal kidney, prostate and lymph
 CC node mRNAs. The NHPs of the invention are all apparently encoded on
 CC chromosome 7 and share sequence similarity with animal kielin proteins.
 CC Kiehlins are secreted proteins implicated in a number of biological
 CC processes such as development and signal transduction. The NHP nucleic
 CC acids and proteins may be used in the diagnosis and treatment of human
 CC kielin-related disorders, and may also be used in drug screening and
 CC clinical trial monitoring. They may further be used in cosmetic and
 CC nutraceutical applications. The present sequence represents an NHP
 CC of the invention.
 XX Sequence 1192 AA;
 SQ
 Alignment Scores: 6.57e-06 Length: 1192
 Pred. No.:

Score: 158.00 Matches: 169
Percent Similarity: 27.63% Conservative: 57
Best Local Similarity: 20.66% Mismatches: 219
Query Match: 4.13% Indels: 374
DB: 24 Gaps: 46

US-09-743-237-3 (1-2134) x ABP97376 (1-1192)

QY	14	GGCGAAGCTCGCGGGGGGACACGCGCGGGGAGCTCCTCGGGAGTACCCCGGAT	73
Db	198	GlyCysAlaPheGlyGlyGlyProSerGlyAlaAspPheProHisProSerAsp	217
QY	74	CCGAG-----AGCTCAGCGCGCTGGAGGACGTCCGGCTCTCGAGG	114
Db	218	ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal	237
QY	115	CCCGCGCGCGCGCG-----CCTGCAACGTGCACTTCTCTGCTCTCGCTGC	159
Db	238	ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys	255
QY	160	TACCCGCGCACCGCAGCGCGGGTGTTCCTCCCTCGGGCG-----	201
Db	256	ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln	275
QY	202	-----CCTGGTCCTCGGAGGAGCTCCACCGCGCGCTCCGATGATC	246
Db	276	GluTyrPheSerProProGlyAsp-----ProCysArgArgCysLeu	289
QY	247	CCAGTTGAATCAAGTAAAGCAGGTGTTACTACTACAGTAATAATCCGGAAGAACAAC	306
Db	290	CysLeuAspGlySerValSer-----CysGlnArgLeuPro	301
QY	307	TTTGCAGAATCTTCTGCTCAGGAATCTGTTCCAAAGTCCCATGTTCCCGAGCACTAGA	366
Db	302	CysProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp	320
QY	367	GGATGCTCCTG-----CTGTTCTCTTAAGAAAGATTC---CAACCCAATGTT	411
Db	321	GlyCysLeuTyrGlnGlyGlyGluPheAlaSerGlyGluArgPheProSerProThr-Al	340
QY	412	GATATGCCAATTAAGGGGGCACACAATGCTATGATATAGCAATCTAGACACAGA	471
Db	340	aAlaCysHisLeu-----CysLeuCysTyrGlnGlySerValSerCysG1	355
QY	472	ACTAAAGCACTC-----CATTTGTTCTCT-----	496
Db	355	uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs	375
QY	497	-----CAGTATCAAGATCAAAATAATTATCTACAGTCAGATGTCCTTAACC	543
Db	375	pCysAspGlyCysGluTyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspPr	395
QY	544	A-----ATGACTCTTTAGTAGGAGATTTTCCAGCATCAACAAAT	588
Db	395	oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys-----	411
QY	589	AAATCTCATTACAAACAACCTTGGAGGAGCTTACCATTGCTAGTCAACGGGTCTGCTT	648
Db	412	-----GlyArgArgProCysGluProProGlyCysSerHi	423
QY	649	CCCC-----	652
Db	423	sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa	443
QY	653	-----TCGGGATCACTCTCCAGGACCAACCAAAATAACTTTGGCTGGTACTG	702
Db	443	lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr-----	458
QY	703	TGACTGCTTGGCAGTGGGACCTTTCGCAACAACCTGCAATTGTAAATAATTTGTGCAACA	762
Db	459	-----CysSerLeuCysThrCysGln-----GlyAr	467
QY	763	CTTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGGCATGCTTGGTAGAAATCCAGA	822
Db	467	gGluHisGlnAspGlyGluGluPhe-----	475
QY	823	AGCTTTCAGCCCAAAATTTGGGAAGGCGCAATTTGGCAATGTCAAGCCCCCAGCACACAA	882
Db	476	-----GluGlyProAlaGlySerCysGlu-----	483
QY	883	AGGTCGAACATGCAGGAGGTTCAGGCTGCCTGAAAGAATTTACTGCGAGTGTCTATGAGGCCCA	942
Db	484	-----TriPysArgCysGlnAlaGlyG1	491
QY	943	AATTATGTGT-----	952
Db	491	nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr	511
QY	953	-TCTTCTATTTCG-----AAATGCATTTGTTGCAAAATTTAT---GAAGAAAGCCAGACAG	1005
Db	511	gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro-----	529
QY	1006	AAAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGGAGGACGACCATTA	1065
Db	530	-----GluGlySerArgTr	534
QY	1066	CCTGCCACCAACGAATTTTCAGGACTTCCAAGATTTCAGTCAGATAGGCGGCTTCCTC	1125
Db	534	pValProPro-----AspSerAlaCysSerSe	543
QY	1126	ATCATCTCTCTGGAGGTGGTGAGGCCACATCGGCC-----TGCCTG-----CT	1170
Db	543	rCysValCysHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCys	562
QY	1171	TGCTCAGGAGAGAAGCGCCGAGAAAGAACACTGCTCAAGTGCCTGCCAGAGCAGATGAT	1230
Db	562	sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys-----	576
QY	1231	CCTGGAGGAATTTGGAAGTGTTCATCATCAGATTTCTCCACACTGAGTTTAAATCTAAGG	1290
Db	577	-----SerAspCysGluHisGluG1	583
QY	1291	ATTGAANTGGAG-----TAGAGTATAAGTGTGAATGCATGTT	1329
Db	583	YArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysI1	603
QY	1330	GATTTTGTCTTAGTCTAGAAATCTAGTTTAGAAAGGATGTTTAGGGAACATGAGGCT	1389
Db	603	e-----	603
QY	1390	GGCTCTGCAGCAACACACAGGCTCCCTCGCATCCCTGGGCCAGGAGGATTTACTCAGAGC	1449
Db	604	---CysGluProGlnProGlyProSerLeu-----	614
QY	1450	TCCTCTGAAGATG---GGCAACCATGCCCTTTCTGAGGA-----	1489
Db	615	---ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe	632
QY	1490	---GGTGCATGCCCTGAGCATTTGT-----	1510
Db	632	uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe	652
QY	1511	-----TTGCTGCGCCCA-----GA	1524
Db	652	rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysG1	672
QY	1525	GGAGAGAGCTTGGTTTCCCAT-----	1546
Db	672	nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerG1	692
QY	1547	-----GTCTGGGAGAGTGTCTGAGGCGCGCGG	1575
Db	692	uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaG	712
QY	1576	AGGCGAGACAGGCCCTCGGAGAGCTCACTCTGGTGGAGCTCTTCTCTCAGAGATGT-	1634

Db 712 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 730
 QY 1635 -----TGCTCTGGAGGC-----TGCTCTGGAGGC-----TGCTCTGGAGGC 1659
 Db 730 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluCysGlnAla---- 748
 QY 1660 AACCTTAATGGTTCCTGTTGTTTCAAAATATTAGAAATAAGTCTCCGGATGGC 1719
 Db 749 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 763
 QY 1720 TGTGTGATACACCTTAAATCTCTAGACACTAGACACTAAAGATTTCTGTAGC 1779
 Db 764 CysCysGluArg-----CysGln 769
 QY 1780 GTAGATATTCCCA-----GAGACACGGAACTCTCAGTCTTCTTCAAGGCCCC 1830
 Db 770 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 784
 QY 1831 GGGAGACGAGCAATGGGGCCTCGCAGCCAGGCTTGACACGATGCTTTGAGTTAGA 1890
 Db 785 GlyGluArg-----TriThrVal-----AspThrCysThrSerCysSerCysMetaAla 800
 QY 1891 GGACTTAAATATCCAGTTCTTCTGCTGTTTCTACTTGAATGTGGA 1938
 Db 801 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 816
 RESULT 15
 ID ABP97377
 XX ABP97377 standard; Protein; 1207 AA.
 XX AC ABP97377;
 XX DT 10-MAY-2003 (first entry)
 XX DE Human kielin-like protein NHP10, SEQ ID NO:20.
 XX KW Human; kielin-like protein; NHP; chromosome 7; development;
 KW signal transduction; kielin-related disorder; diagnosis; treatment;
 KW drug screening; clinical trial monitoring; cosmetic; nutraceutical.
 OS Homo sapiens.
 XX WO2003004609-A2.
 XX 16-JAN-2003.
 XX 03-JUL-2002; 2002WO-US21118.
 XX 03-JUL-2001; 2001US-302949P.
 XX 29-AUG-2001; 2001US-315634P.
 XX (LEXI-) LEXICON GENETICS INC.
 XX Walke DW, Scoville J, Turner CA;
 XX WPI; 2003-210352/20.
 XX N-PSDB; AB275048.
 XX New human kielin-like nucleic acids, useful for preparing a composition
 PT for treating human kielin-related disorders -
 XX Claim 3; Page 75-77; 85pp; English.
 CC The invention relates to novel human kielin-like proteins (referred to as
 CC NHPs; ABP97368-ABP97379) and cDNAs encoding them (AB275039-AB275050). The
 CC invention also encompasses recombinant expression vectors and host cells
 CC comprising a nucleic acid of the invention. The NHP cDNAs were obtained
 CC by aligning cDNAs from human kidney, foetal kidney, prostate and lymph
 CC node mRNAs. The NHPs of the invention are all apparently encoded on
 CC chromosome 7 and share sequence similarity with animal kielin proteins.
 CC Kiehlins are secreted proteins implicated in a number of biological
 CC processes such as development and signal transduction. The NHP nucleic
 CC acids and proteins may be used in the diagnosis and treatment of human

CC kielin-related disorders, and may also be used in drug screening and
 CC clinical trial monitoring. They may further be used in cosmetic and
 CC nutraceutical applications. The present sequence represents an NHP
 CC of the invention.
 XX SQ Sequence 1207 AA;
 Alignment Scores:
 Pred. No.: 6.6e-06 Length: 1207
 Score: 158.00 Matches: 169
 Percent Similarity: 27.63% Conservative: 57
 Best Local Similarity: 20.66% Mismatches: 219
 Query Match: 4.13% Indels: 374
 DB: 24 Gaps: 46
 US-09-743-237-3 (1-2134) x ABP97377 (1-1207)
 QY 14 GCGCAAGCTCGCGGGGCGCAGCAGCGCGGGAGCTCTCGGGAGTACCCCGGAT 73
 Db 213 GlyCysAlaPheGlyGlyLysGluTrpProSerGlyAlaAspPheProHisProSerAsp 232
 QY 74 CCCAG-----AGCTCAGCGCGCTGGAGGACGCTCGCGCTCCCTCAGG 114
 Db 233 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgArgCysVal 252
 QY 115 CCCCAGCGCGCCG-----CCTGCAACGTGCATCTCTCTCTCTCTCTCTCT 159
 Db 253 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 270
 QY 160 TACCGCGCGCAGCGCGCGGGTGTGTTGCGCCCTGGGGCG----- 201
 Db 271 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 290
 QY 202 -----CCTGGGTCTCGCGAAGGAGCCTCCCGCGCGCTCCGATGATC 246
 Db 291 GluTrpPheSerProProGlyAsp-----ProCysArgArgCysLeu 304
 QY 247 CCAGTGTAAATCAAGTAAAGTAAAGTAAAGTAAATCAAGTAAATCAAGTAAATCAAG 306
 Db 305 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 316
 QY 307 TTTGCAAGTCTCTCTCTCAGCAATCTCTGTTGCAAGTCTCTCTCTCTCTCTCTCTCT 366
 Db 317 CysProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 335
 QY 367 GGATGCTCTCTG-----CTGTTCTCTTAAAGAAAGATTC---CAACCCATGCT 411
 Db 336 GlyCysLeuTrpGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 355
 QY 412 GATATGCCAATTTAAAGGGGCGACAAATGCTATGATATGATATGATATGATATGATATG 471
 Db 355 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGln 370
 QY 472 ACTAAAGCAGCTC-----CATTTGTTCTCT----- 496
 Db 370 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 390
 QY 497 -----CAGTATCAAGATCAAAATAATTATCTACAGTCAGATGTCCTCAAAACC 543
 Db 390 pCysAspGlyCysGluTrpLeuGlyCysTrpLeuSerAsnGlnGluPheProAspPr 410
 QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTCCAGCATCAACAAATTT 588
 Db 410 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 426
 QY 589 AAATCTCATTAACAACAACACTTGGAGGAGCCCTTACATCGGTAGTACGAGGGTCTGCTTT 648
 Db 427 -----GlyArgArgProCysGluProProGlyCysSerHi 438
 QY 649 CCCC----- 652
 Db 438 sProLeuLeuProSerGlyHisCysCysProThrCysGlnGlyCysArgTrpHisGlyVa 458


```
QY 653 -----TCGGATCAACTCTCCAGSACCACCAAAATAACTTTGGCTGGGTACTG 702
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
458 lThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 473
QY 703 TGACTGCTTTGCCAGTGGGACTTTTGGCAACAACATGTAATTAATTTGTGCAACAA 762
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
474 -----CysSerLeuCysThrCysGln-----GlyAr 482
QY 763 CTTGCATCATGATTAAGCGTTTAAAGCCATTAAAGCATGCTGTTAGTAATCCAGA 822
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
482 gGluHisGlnAspGlyGluPhe----- 490
QY 823 AGCTTTCCAGCCCAAAATTTGGGAAGGCCAATTTGGCAATGTCAAGCCCGCAGCAACAA 882
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
491 -----GluGlyProAlaGlySerCysGlu----- 498
QY 883 AGGTCGCAACTGCAGGAGTGCAGCTGCCTGAAGAATTACTGCGAGTGTATGAGGCCCA 942
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
499 -----TrpCysArgCysGlnAlaGlyGI 506
QY 943 AATTATGTGT----- 952
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
506 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 526
QY 953 -TCTTCTATTTCG---AAATGCTATTGTTGCAAAAATTAT---GAAGAAGCCCGCAGAACG 1005
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
526 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 544
QY 1006 AAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGTTTGGAGGCCACCATTA 1065
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 -----GluGlySerArgtr 549
QY 1066 CTGCCACCAACGAAATTTTCAGGACTTCCAGATTTCAGTCAGATAGCGGCGCTTCCTC 1125
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 pValProPro-----AspSerAlaCysSerSe 558
QY 1126 ATGCATCTCTGGGAGTGGTGGAGGCCACATCGGCC-----TGCTG-----CT 1170
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
558 rCysValCysHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 577
QY 1171 TGCTCAGGAGAGAGCCGCGAGAAAGAACACACTCTCCCAAGTGCCTGCAGGAGCATGAT 1230
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
577 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 591
QY 1231 CTTGGAGGAATTTGGAAGTGCCTATACAGATTCTCCACAGTTTAAATCTAAGGG 1290
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
592 -----SerAspCysGluHisGluGI 598
QY 1291 ATTGAAATGGAG-----TAGAGTATAAGTGTGAATGCATGTT 1329
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
598 yArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysII 618
QY 1330 GATTTGTCTTAGTCTAGAAATCTTAGTTTAAAGAGTGTTTAGGGGAACATGAGGCT 1389
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
618 e----- 618
QY 1390 GGCTCTGCAGCAACAACAGGCTCCCTGCATCCCTGGGCCAGGAGTTTACTCAGAGC 1449
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
619 -----CysGluProGlnProGluGlyProProSerLeu----- 629
QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCTTTTCTGAGGA----- 1489
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
630 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 647
QY 1490 -----GGTGCATGGCCTGAGCATTTGT----- 1510
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
647 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 667
QY 1511 -----TTGCTGGCCCA-----GA 1524
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
667 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGI 687
QY 1525 GGAGAGAGCTTGGGTTCCCAT----- 1546
```

```
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
687 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGI 707
QY 1547 -----GTCCTGGGAGAGTGTCTGCAGGCGCGCG 1575
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
707 uArgHisThrProGlySerCysCysProValCysArgGluCysValVal-GluAlaG 727
QY 1576 AGGCGAGAGCAGGCCCTGCGGAGAGCTCACTCTGGTGCAGACTCTTCTCTCAGAGAATGT- 1634
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
727 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 745
QY 1635 -----TGCTCTGGAGGC-----TGCTCTGCATGAA 1659
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
745 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla---- 763
QY 1660 AACCTTAATGTTTCTGTTGTTTTCAAATTTAGAAATAAGTTCTCCGATGGGC 1719
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 778
QY 1720 TGTGTGATACCACTTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTTCTGTAGC 1779
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
779 CysCysGluArg-----CysGln 784
QY 1780 GTAGATATTTCCTCCA-----GAGACACGGCAACTGTCACTTTCTTAAGGCCCCC 1830
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
785 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 799
QY 1831 GGGAGAGCGCAGGCAATGGGCGCTCGCAGGCCAGGCTTGCACACCATGTTCTTGAGTTAGA 1890
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
800 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMetaIa 815
QY 1891 GGACTTAAATATTATCCAGTTTCTTCTGTGTTTCTTACTTGAATTGTGGA 1938
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
816 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 831
```

Search completed: July 24, 2003, 13:58:40

Job time : 129.554 secs

